Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Learning with Reduced Feedback Information: M. W. SHELLY 209
Ball Throwing Responses to Photographically Portrayed Targets: P. C. SMITH AND O. W. SMITH 223
Effects of Amount and Percentage of Reinforcement and Number of Acquisition Trials on Conditioning and Extinction: A. R. WAGNER 234
Stimulus Attributes and Drive in Paired-Associate Learning: H. LEVITT AND A. E. GOSS 243
The Observing Response in Discrimination Learning: R. C. ATKINSON 253
Interaction of Lack of Sleep with Knowledge of Results, Repeated Testing, and Individual Differences: R. T. WILKINSON 263
Figure Coherence in the Kinetic Depth Effect: B. F. GREEN, JR. 272
Concept Identification Using Simultaneous Auditory and Visual Signals: D. S. LORDAHL 283
Eyelid Conditioning Performance as a Function of Emotion-Producing Instructions: K. W. SPENCE AND H. GOLDSTEIN 291
Proactive Inhibition of Connected Discourse: N. J. SLAMECKA 295
Retention of Information under Conditions Approaching a Steady State: R. N. SHEPARD AND M. TEGHTSOONIAN 302
Supplementary Report: Information Deprivation and Irrelevant Drive as Determiners of an Instrumental Response: A. JONES 310
Supplementary Report: The Influence of One Stimulus on the Prediction of the Alternative Stimulus in Two-Choice Problems: I. RUBINSTEIN 311

American Psychological Association

Vol. 62, No. 3 September 1961

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Production Manager

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 62, NO. 3 SEPTEMBER 1961

LEARNING WITH REDUCED FEEDBACK INFORMATION [1]

MAYNARD W. SHELLY

Office of Naval Research

The information *S* receives concerning his responses undoubtedly affects the rate at which he learns the correct response. Although this is obvious, two problems remain largely unexplored. These are the relationship between amount of feedback information and initial rate of learning, and the relationship between amount of feedback information and the general shape of the empirical learning curve.

Most investigations on the effects of knowledge of results have simply varied the feedback information through eliminating or reducing the number of events following *S*'s response. This is true of the old line drawing experiments (e.g., Trowbridge & Cason, 1932) and the more recent studies involving contingent and noncontingent feedback (e.g., Bourne & Pendleton, 1958; Neimark, 1956; Shelly, 1960). Feedback information giving the correctness of a response may also be decreased by increasing the number of events following a response, which may be considered to be noise added to the information about the correct response. It is the purpose of the present study to investigate some of the effects of adding such noise to the information fed back to *S* within a paired-associate paradigm.

This paper consists of three experiments, each of which varied some characteristic of the noise added to the information given *S* about the correct response. Experiment I varied the amount of noise which was added to the correct information. In Exp. II the sequential characteristics of the noise over trials was varied, while in Exp. III the distribution of noisy events over the set of available incorrect response alternatives was varied.

Each experiment asked two questions. The first question was whether the reduction in initial learning rate was proportional to the amount by which the noise decreased the infor-

[1] The data for these experiments were collected while the author was at the University of Wisconsin. Some of the results together with some results not included in this paper were presented at the Midwestern Psychological Association meeting in St. Louis in 1956. The author wishes to thank Northwestern University for providing him with facilities to complete the analyses of these experiments, the Office of Naval Research for encouraging the professional interests of its scientists, and Carolyn Shelly for helping in the editing of the final draft of the paper.

mation fed back to *S*. (Here, and in the following, a decrease in information refers to a decrease in the knowledge needed to identify the correct response, rather than to a decrease in uncertainty.) The second question was whether the reduction in information produced the same temporal pattern of effects for all conditions. This latter question is part of the question as to whether difficulty of the task alters the shape of the learning curve.

General Method

Task.—The task employed in these experiments was a slight modification of the usual paired-associate paradigm. The *S* was to learn to associate responses (pulling switches) with visual stimuli. All conditions employed eight stimuli. The stimuli were either angles or modifications of a triangle. In some conditions there were four response alternatives and in others, eight alternatives were available. When there were eight response alternatives each stimulus had a unique response associated with it; when there were only four response alternatives, each response alternative was correct for exactly two stimuli. The stimuli were presented to *S* on a screen placed before him. Whenever a stimulus appeared on the screen *S* was to attempt to select that response alternative which would be correct for the stimulus being presented.

Under conditions with no noise, a lamp would light following the response made by *S*; this lamp, located above some switch, indicated which switch was correct for the preceding stimulus. Under the conditions with noise, more than one lamp would light following a response by *S*. The additional lamp (or lamps) represent what is referred to as noise in the experiments in this paper. Only one of the lamps, however, would *always* come on for a given stimulus; this lamp, the constant lamp, would indicate the correct response for the given stimulus. The other lamps, occasionally lit after a response, will be called noise lamps. Thus, in the conditions with no noise, *S* was required only to associate the correct response with each stimulus. In the conditions with noise, *S* had both to identify the correct response for a given stimulus and associate this response with the stimulus, although these two tasks are not really distinct.

Apparatus.—The apparatus consisted of two parts; an experimental booth in which *S* sat, and the programing and recording equipment exterior to the booth. The interior of the booth was flat black. On a sloping table in the booth, built into an aluminum chassis, were the response switches and the green-domed feedback lamps mounted above their corresponding switches. Those switches not in use were covered by a box. The stimuli were presented by strip film projector, and the feedback programed on Western Union Tape Transmitters.

Subjects.—The *S*s were volunteers from elementary psychology courses at the University of Wisconsin. Each group was approximately equally divided between men and women.

General procedure.—When *S* reported to the laboratory he was, within the restriction that the same number of *S*s serve in each condition, randomly assigned to one of the conditions of the experiment.

In the instructions *S* was told that this was a learning experiment. He was told that one switch would be correct for each different stimulus. He was to learn to pull the correct switch each time another stimulus appeared. If he were assigned to a condition free of noise, he was told that the lamp which came on after he made a response would indicate the correct response for the preceding stimulus. If *S* were in a noisy condition he was told that several lamps would come on after his response, but that only *one* of these lamps would *always* light at the end of a trial; the switch under this lamp was the correct switch. His task, he was told, was to learn which lamp would follow each stimulus just as in everyday life we learn to anticipate consequences which *always* follow a given action even though on any one occasion, many things might happen. A brief demonstration was then given.

The stimuli were presented to *S* at a 4-sec. rate. About .25 sec. before the stimulus changed one or more of the green-domed lamps came on and stayed on until the stimulus changed.

In Exp. I the stimuli were presented in blocks of 80 trials; blocks of 96 trials were used in Exp. II and III. In each block of trials each stimulus occurred equally often (10 times in Exp. I and 12 times in Exp. II and III); the stimuli were assigned their sequential positions randomly without replacement. Three different blocks of stimuli were employed and each block was presented twice. The six blocks composing each experiment were presented in a random order.

Notation for and general characteristics of the experimental conditions.—To make it possible to describe the conditions of this paper succinctly, the following notation will be used. A parenthesis containing either four or eight numbers will be used to represent each experimental condition. Assume reference is being made to a single fixed stimulus. Then, the first position in the parentheses will be reserved for information concerning the response alternative which is correct for this fixed stimulus; the number in this first position will equal the number of times this given stimulus appears in a block of trials. The lamp associated with this position is the constant lamp. The remaining positions denote the number of times a noise lamp will become lit above another alternative following *S*'s responses.

As an example of this notation, in Exp. I there were three conditions: in the first condition no noise occurred, in the second condition exactly one noise lamp occurred, and in the third there were two noise lamps. Then the sum of all numbers in the parentheses beyond the first number will equal zero if there is no noise, will equal the number of times the stimulus appears in a block of trials if one noise lamp occurs with each stimulus, or will equal twice the number of times the stimulus appears in a block of trials if two noise lamps occurred with each stimulus. Thus if the stimulus appears 10 times in a block of trials, possible admissable conditions would be the following: (10, 0, 0, 0), (10, 0, 0, 0, 0, 0, 0, 0), (10, 2, 4, 4), (10, 4, 8, 8), (10, 2, 4, 4, 0, 0, 0, 0,), (10, 2, 2, 2, 2, 2, 0, 0), and (10, 8, 8, 1, 1, 1, 1, 0). The first two representations would be for conditions free of feedback noise having four and eight available alternatives, respectively. The numbers included in the third parentheses represent a condition in which a false or misleading lamp comes on above one wrong alternative twice and above the other two incorrect alternatives four times in a block of 10 trials. To control against biases arising from the placement of alternatives, especially in Exp. II and III, two different associations of correct stimuli with response switches were used for each condition.

Measure of information.—Since the purpose is to investigate some of the effects of decreasing information in a learning situation, it is necessary to employ some measure of information. Except for Exp. II, where an additional measure of information is used, the measure of information was computed as follows. For each stimulus in the sequence of stimuli presented to *S*, determine the expected (average) number of times this stimulus must be presented to an *S* with perfect memory before he can identify the correct switch for that stimulus. Take the logarithm to the base two of the number of response alternatives available to *S*. Divide this number by the expected number of trials required to identify the correct switch. This value will be designated by an *I*. *I* values can be viewed roughly as the average number of bits presented to *S* per trial. When the number of alternatives is fixed, this value is simply the reciprocal of number of trials required to identify the correct alternative for the fixed stimulus.

Experiment I

The purpose of Exp. I was to study some effects of the amount of noise upon learning through varying the number of noisy events which followed a response.

Method

Conditions.—There were six conditions. In the following list of conditions, the values in the parentheses, in other than the first position, represent expected values over three blocks of trials:

1. (10, 0, 0, 0), $I = 2.00$
2. (10, 3.3, 3.3, 3.3), $I = .86$
3. (10, 6.7, 6.7, 6.7), $I = .47$
4. (10, 0, 0, 0, 0, 0, 0, 0), $I = 3.00$
5. (10, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4, 1.4), $I = 1.43$
6. (10, 2.9, 2.9, 2.9, 2.9, 2.9, 2.9, 2.9), $I = .97$

Conditions 1 and 4 were conditions free of noise. In Cond. 2 and 5 one noise lamp, occupying a position selected randomly from among the incorrect positions, appeared on each trial. In Cond. 3 and 6 two noise lamps appeared on each trial; the sequence of pairs of positions occupied by the noise lamps was random in these two conditions.

Subjects and stimuli.—A total of 10 *S*s served in each condition. Four of the *S*s were, however, replacements for *S*s who apparently could not understand the instructions. These 4 *S*s, all in noise conditions, showed no signs of learning anything, and at the end of the experiment could not explain what they were supposed to have done in the experiment or gave erroneous explanations.

The stimuli used were eight angles: 0°, 45°, 90°, 135°, 180°, 225°, and 315°. All were drawn in standard position.

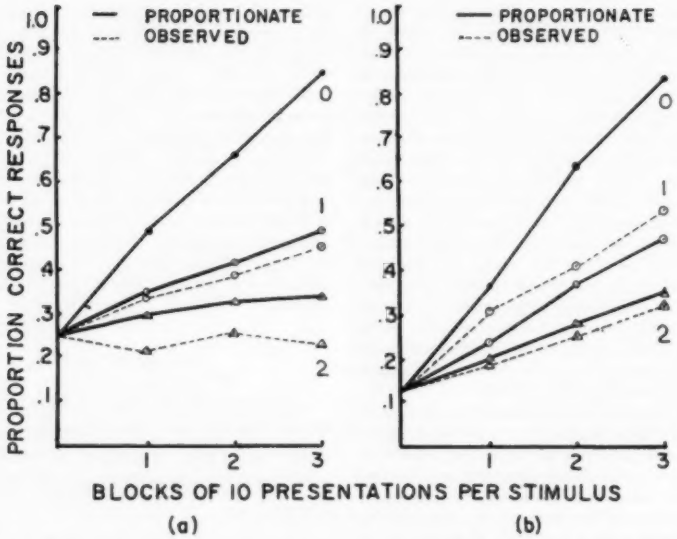

FIG. 1. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials in Exp. I for conditions with four-alternative (a) and eight-alternative (b) responses.

Results

Initial rate of learning.—To examine the effects of reduced feedback information upon initial rate of learning only the data from the first three blocks of trials were used. After the first three blocks, the no-noise groups were approaching perfect performance. This made it difficult to determine the appropriate predicted performance for the noise groups.

In Fig. 1 are presented the data for the first three blocks of trials. The curves in Part a are for four-alternative conditions; those in Part b for the eight-alternative conditions. The curves in Parts a and b labeled "O" are the learning curves for the conditions free of noise. These curves act as both empirical results and the basis for the theoretical predictions of proportionality. Below these curves are both solid and broken curves. The solid line curves below these O curves are computed on the basis of *I* values. They represent the percentage of correct responses which would result had *S*s shown a loss, for a fixed block of trials, proportionate to the decrease in information. That is, they were computed as follows. The value on the O curve was taken for a fixed block of trials. This was then multiplied by the ratio of the *I* value for the curve being computed to the *I* value for the appropriate O curve. The dotted curves represent the obtained results. Data points for each pair of predicted and obtained curves are enclosed by identical symbols. Each set of solid and broken curves are labeled by a number. The numbers represent the number of noise lamps which appear per trial in the conditions upon which the curves are based. As can be observed for both four (Part a) and

eight (Part b) alternatives, the correspondence is generally quite fair. For the condition with four alternative responses, the prediction of proportionality agrees with the results within 19.4% error. For the eight-alternative response conditions, the agreement is better, to within 14.0% error.

To obtain a measure of the correspondence between the proportionate predicted values and the obtained values, the following measure was employed. Let X_i be a random variable, normally distributed with mean zero and variance 1. Then define the derived random variable W to be:

$$W = \sum_{i=1}^{N} \tfrac{1}{2}(X_i + X_{i+1}) - c(X_i, X_{i+1}) \frac{X_i X_{i+1}}{X_i + X_{i+1}}$$

where $c(X_i, X_{i+1}) = 0$ if sign (X_i) = sign (X_{i+1}), i.e., both X_i and X_{i+1} have negative values or both variables have positive values; or $c(X_i, X_{i+1}) = 1$ if sign $(X_i) \neq$ sign (X_{i+1}). W then represents the random variable which takes on as values the *area* between the predicted and obtained curves, when the distances between corresponding points on the curves have been transformed into appropriate scores based on standard deviations and successive points along the abscissa are separated by unit distances. The area obtained in this manner will be called a *normalized deviation area*.

W was used as a measure of the amount of correspondence between the predicted and obtained results. This measure was used in preference to others for two reasons. First, this measure takes into account the amount of variability at each data point; small deviations between curves can be quite important if the variability is correspondingly small. Secondly, the use of areas instead of distances between a small number of corresponding points, is sensitive to the sequential ordering of data points about the theoretical points. If the data points lie consistently to one side of the theoretical points, then the normalized deviation areas would be larger than were the two curves to cross several times.

For Experiment I, the normalized deviation areas are 1.97, 8.68, 1.68, and 2.10 for Cond. 2, 3, 5, and 6, respectively.

Shape of the learning curve.—The following analysis is directed towards the effects of reduced information upon the shape of the performance curve. There are many specific questions which could be asked on the general problem of shape of learning curves. One of the most interesting is the location, in the empirical learning curve, of the point at which the greatest increase in the number of correct responses takes place. This of course has a direct implication for the shape of the learning curve. For the usual learning curve, this place of maximum increase would be between the beginning of the experiment and the end of the first block of trials. If it occurred elsewhere, the learning curve would not be monotonically negatively accelerated. The following analysis will be concerned with the maximum increase in the proportion of correct responses between successive blocks of trials; this quantity will be denoted by MaxΔR. The chief interest will be the location of the MaxΔR in the learning sequence. In Fig. 2 the value of MaxΔR is given by the height of the bars. It may be seen that the place at which MaxΔR occurs varies with experimental conditions. The meaning of "Block after which MaxΔR occurs" in Fig. 2 is as follows: If the maximum increase in the proportion of correct responses occurs between (*a*) the proportion of correct responses which could be expected on a chance basis and (*b*) the proportion of correct responses obtained on the first block of trials, MaxΔR is considered to have occurred after block "O." If MaxΔR occurs between the second and third block of trials, it is included in the graph as occurring after Block 2. The last place MaxΔR could occur in this experiment would be between Blocks 5 and 6, or after

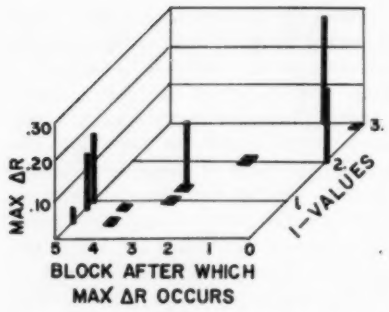

FIG. 2. Amplitude of MaxΔR, and the relation between the block after which it occurs and *I* value of the feedback information for Exp. I.

Block 5 in Fig. 2. This dimension located the *position* of MaxΔR in the learning sequence. The other dimension on the floor of the graph simply gives the *I* values listed previously for the conditions of Exp. I.

Sometimes the differences between the increases in proportions of correct responses between blocks are quite small. It is consequently desirable to establish the reasonableness of two types of statistical hypotheses. First, it is desirable to establish that the distribution of positions of MaxΔR cannot be accounted for by chance deviations from either the zero or first block positions. Secondly, because the relationship of the position of the MaxΔR values to the *I* values is important, it is also desirable to obtain some estimate of this relationship.

The probability of finding a set of deviations in the position of MaxΔR from the zero block position as large or larger than those obtained in this experiment, under the hypothesis of a rectangular distribution, will be denoted by P_0. The probability of finding a similar set of deviations from the Block 1 position as large or larger than those obtained will be denoted by P_1. For this experiment the values of these two probabilities were:

$$P_0 = \frac{15}{6^5} < .001$$

and

$$P_1 = \frac{12}{5^5} < .004$$

It is thus unlikely that MaxΔR occurs during the first part of the learning sequence in all conditions.

Consider a line drawn from the corner of the zero block and *I* value of 3.00 to the diagonally opposite corner, Block 5 and *I* value zero; this line would result from a perfect linear relationship between MaxΔR and *I* value which passed through the two described corners. This line has no theoretical significance. It was chosen because if the relationship between position of MaxΔR and the *I* value approximates this line, it will be clear that the "shape of the learning curve" depends upon the *I* value of the feedback information. The positions that the columns would occupy were there such a perfect relationship in Exp. I between these two variables are indicated by the rectangles on the floor of the graph; these rectangles do not lie solely on the diagonal. The rectangles in Fig. 2 could lie solely on the diagonal for the *I* values used only if MaxΔR could assume values other than whole numbers. Since this is not logically possible within the framework of the experimental analysis, the illustrative rectangles are placed on the *nearest* whole number of the dimension "Block after which MaxΔR occurs." This results in their only approximating a diagonal.

The probability of finding a set of deviations from this hypothetical diagonal line as small or smaller than those obtained, under the hypothesis of a rectangular distribution, will be denoted by $P_{Diag.}$. This gives the proba-

ability that, as graphed, there is not a relationship close to a linear relationship passing through the two corners between the position of MaxΔR and the I values of the conditions of Exp. I. The value of this probability was: $P_{Diag.} = \frac{2}{6^4} < .002$. Therefore the probability that the position of MaxΔR is not closely related to the I values of feedback is small. Consequently, the shape of the learning curve varies with information fed back.

It will be noted that although there seems to be some tendency for the values of MaxΔR to decrease as the I values of the feedback decrease, the size of the decrease is not striking. The tau correlation coefficient between these two variables is, however, .866.

Experiment II

In Exp. I the I values of the feedback information were changed by varying the number of noisy events which followed a response. Experiment II attempts to extend these results to conditions in which the I values are changed by varying the sequential patterning of the noisy events. Patterning is varied principally by varying the number of times a noise lamp remains in the same position for successive presentations of a given stimulus.

Method

Conditions.—The conditions for Exp. II are presented in Table 1, together with the I values for each condition. For all conditions, except the two noise-free groups, two lamps came on following each S's responses. One of these lamps, the constant lamp, indicated the correct response; the other was a noise lamp. In all but one of the noise conditions, (Random$_1$), each stimulus was paired with two positions which the noise lamp could occupy. A lamp in one of these two positions would light, together with the constant lamp indicating the correct response alternative, following each presentation of the appropriate stimulus. The effects of the sequential patterning of the position of noise lamps are the particular interest in this experiment.

TABLE 1

Conditions of Exp. II

Conditions: Frequency or Pattern of Change in Noise Lamp	4-Alternative Conditions		8-Alternative Conditions	
	I_1	I_2	I_1	I_2
No noise (N)[a]	2.00	2.00	3.00	3.00
Alternating (A)	1.00	1.00	1.50	1.50
Random$_1$ (R$_1$)	.78	.78	1.43	1.43
Random$_2$ (R$_2$)	.67	.67	1.00	1.00
By quarter blocks (Q)	.72	.50	1.09	.67
By half blocks (H)	.58	.29	.87	.43

[a] The letters in parentheses denote the symbols by which the conditions are referred to in Fig. 3 and 4.

In the alternating noise conditions (Cond. A), the two noise lamp positions for each stimulus are employed alternatively. After the presentation of a given stimulus a noise lamp in one position lights; on the next presentation of this stimulus the noise lamp in the other position comes on. The position of the noise lamp continues to alternate in this fashion.

In the conditions designated as "By Quarter Blocks" (Cond. Q), one noise lamp position was employed successively for three presentations of a given stimulus; then the noise lamp in the other of the two positions it assumed would come on for the next three presentations of the same stimulus. Thus the noise lamp position which was paired with a given stimulus changed every quarter of a block.

In those conditions labeled "By Half Block" (Cond. H), the noise lamp position changed every half block. The noise light would come on in one position for six presentations of a given stimulus and then the other position of the noise lamp would be employed for the next six presentations remaining in the block. Thus, the noise lamp changed positions at the beginning of a block of trials, and once more half way through the block.

The no-noise conditions are the same as Cond. 1 and 4 in Exp. I, except that each stimulus is presented 12 times in one block.

Two types of random conditions were used in this experiment. Random$_1$, (Cond. [12, 4, 4, 4] and [12, 1.7, 1.7, 1.7, 1.7, 1.7, 1.7, 1.7]), was the same as Cond. 2 and 5 in Exp.

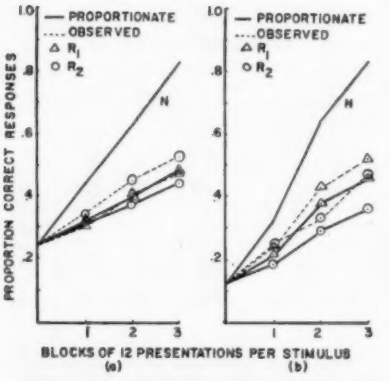

FIG. 3. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials for the two random conditions of Exp. II. (Four-alternative, a, and eight-alternative, b, conditions.)

I with the exception that the number of stimuli per block was changed. $Random_2$ (Cond. [12, 6, 6, 0] and [12, 6, 6, 0, 0, 0, 0, 0]) acts more closely as a control for conditions of Exp. II in that the noise lamp can assume only one of two incorrect positions. For $Random_2$ conditions the probability on any one trial of the noise lamp changing from its previous position is .5.

Information measures.—In Table 1, following each condition, two I values are given. I_1 is the same information measure as described in the General Method section. It is based on the expected number of trials required for an S with perfect memory to identify the correct response alternative, when the beginning point in the sequence is chosen at random. I_2 is identical except that the beginning points in the sequence are considered to occur immediately after a change in the position of a noise lamp. This increases the expected number of trials required to identify the correct response alternative for a fixed stimulus in the Cond. Q and H, and since in Cond. A the position of the noise lamp alternates between the two possible positions, I_1 equals I_2. I_2 is employed in addition to I_1 since it is conceivable that an S's efforts to filter out the noise may begin with or be intensified by the occurrence of a change in the noise lamp's position. There are then two bases by which to compare performance in the noise conditions.

Subjects and stimuli.—Ten *Ss* served in each condition. Two of these *Ss* were replacements for *Ss* who apparently did not understand the instructions. The same criteria for rejection were used as in Exp. I.

The eight stimuli used were clearly discriminable modifications of a solid black isosceles triangle. These stimuli were formed with all possible combinations of the following: the presence or absence of a hole in the center of the triangle; the presence or absence of a cut perpendicular to the base which eliminated one corner; and the presence or absence of a cut oblique to the base which eliminated the other corner at the base of the triangle.

Results

Initial rate of learning.—In Fig. 3 are presented the curves which are predicted on the basis of proportionality and the empirical curves obtained for the first three blocks of trials for the two types of random noise conditions. Part a presents the data for the groups with four response alternatives and Part b for eight alternatives. It can be seen that in general the prediction of a decrease in initial learning rate proportional to the decrease in I value is more accurate for the Cond. $Random_1$ than for Cond. $Random_2$. The percentage error between predicted and observed is 11.7 for $Random_1$, while for $Random_2$ it is 21.3%. That the prediction of proportionality is more adequate for $Random_1$ is also apparent in Table 2 which presents the normalized deviation areas. Since there was a difference between $Random_1$ and $Random_2$ the results to be presented below concerning the sequential patterning of the noise lamp positions should possibly be considered to be limited to conditions in which the noise lamp is restricted to two positions. (The findings of Exp. III will further support the suggestion that the number of incorrect positions a noise lamp can occupy must be considered when evaluating the prediction of proportionality.)

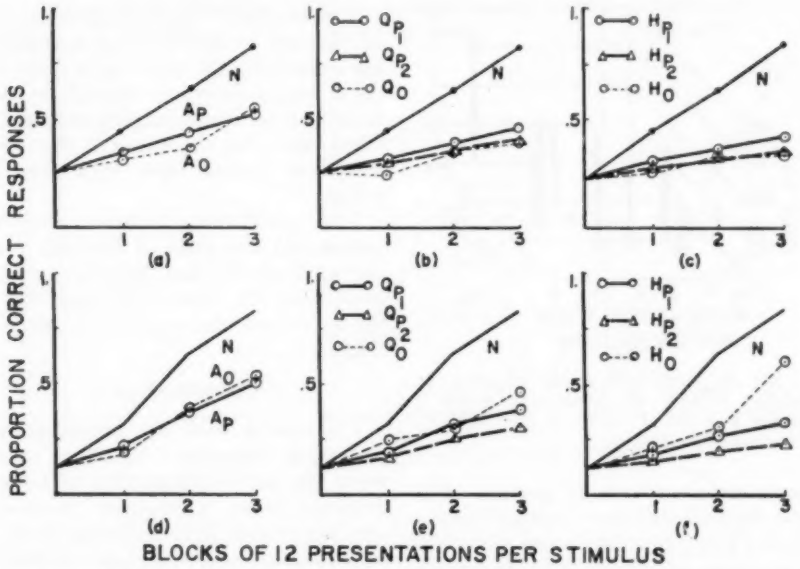

FIG. 4. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials for Cond. A, Q, and H of Exp. II. (For Cond. Q and H, proportionate predictions are made both on the basis of I_1 and I_2. The upper three graphs are for the four-alternative conditions and the lower three for the eight-alternative conditions.)

In Fig. 4 the predicted and obtained curves for Cond. A (Parts a and d), Cond. Q (Parts b and e), and Cond. H (Parts c and f) are presented; the data for the groups with four response alternatives are presented in the upper three graphs. An A, Q, or H with a subscript O refers to the obtained results. The same letters with a P subscript refer to the predicted proportionate results; P_1 indicates that the curve was computed using I_1 as a measure of feedback information, while P_2 indicated that I_2 was used. Curves labeled "N" are based on a noise-free condition.

In general it can be observed that the use of I_2 provides better predictions for the conditions with four response alternatives while predictions based upon the I_1 values are superior for the eight-response alternative conditions. This is generally borne out by the normalized deviation areas presented in Table 2, where the normalized deviation areas involving I_2 are included in parentheses.

The percentage error between the proportionality prediction and the obtained results for the alternating conditions was 7.3, for the four- and eight-alternative response conditions combined. For Cond. Q and H, this

TABLE 2

NORMALIZED DEVIATION AREAS FOR EXP. II

Noise Distribution	4-Alternative Conditions	8-Alternative Conditions
Alternating	2.14	.76
$Random_1$	1.51	2.39
$Random_2$	4.42	3.52
Quarterly	4.53 (1.99)[a]	3.14 (7.58)
Halves	3.44 (.85)	4.50 (7.85)

[a] The values in parentheses represent the normalized deviation areas when I_2 values are used as the standard instead of I_1 values.

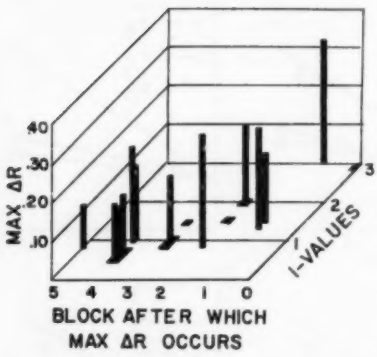

FIG. 5. Amplitude of MaxΔR, and the relation between the block after which it occurs and the I value of the feedback information for Exp. II.

same error was usually greater. For the four-alternative response conditions, when I_1 values were used, the percentage error between predictions of proportionate decrease and the obtained results was 15.3 for Cond. Q and 16.4 for Cond. H; when I_2 values were used the same errors were 7.1% and 4.7%, respectively. The corresponding values for the eight-alternative response conditions for Cond. Q and H, in the same order as above, were 17.3, 36.4, 34.3, and 95.3%.

Shape of the learning curve.—The relationship between the place of occurrence and value of MaxΔR is presented in Fig. 5 with I_1 values used to describe feedback information. The organization of this figure is identical to that of Fig. 2. The following probabilities were obtained in analyzing the data described in Fig. 5:

$$P_0 < \frac{1}{6^5} < .001$$

$$P_1 < \frac{1}{5^5} < .001$$

$$P_{Disc.} < \frac{1}{6^9} < .001$$

The results concerning the location of MaxΔR in the learning sequence are essentially the same as in Exp. I. As I values decrease, MaxΔR tends to occur later in the learning sequence. Thus, here also the shape of the performance curves vary with the I values.

The relationship between the I values and the value of MaxΔR was also analyzed by tau. When I_1 values were used the correlation was .515; for I_2 values, the correlation fell to .273.

EXPERIMENT III

The results from the immediately preceding experiment indicated that when the sequential pattern of noise deviated from a random pattern, the relationship between initial learning rate and the I values of the feedback was affected. In general less proportionality was evident. Experiment III is an investigation of the effects of various nonrectangular *random* distributions of noise lamp positions. It examines the prediction that the inverse proportionality between initial learning rate and amount of noise will be invariant over variations in the proportion of times a noise lamp is associated with a single position. In Exp. II the temporal distribution over trials of the noise lamps was altered; in Exp. III the spatial distribution of noise lamps is varied.

Method

Conditions.—This experiment consisted of 15 conditions, 2 of which were free of noise. The following are the conditions and their I values:

1. (12, 0, 0, 0), I = 2.00
2. (12, 0, 0, 0, 0, 0, 0, 0), I = 3.00
3. (12, 4, 4, 4), I = .86
4. (12, 4, 4, 4, 0, 0, 0, 0), I = 1.29
5. (12, 2, 2, 2, 2, 2, 2, 0), I = 1.36
6. (12, 6, 6, 0), I = .67
7. (12, 6, 6, 0, 0, 0, 0, 0), I = 1.00
8. (12, 6, 3, 3), I = .75

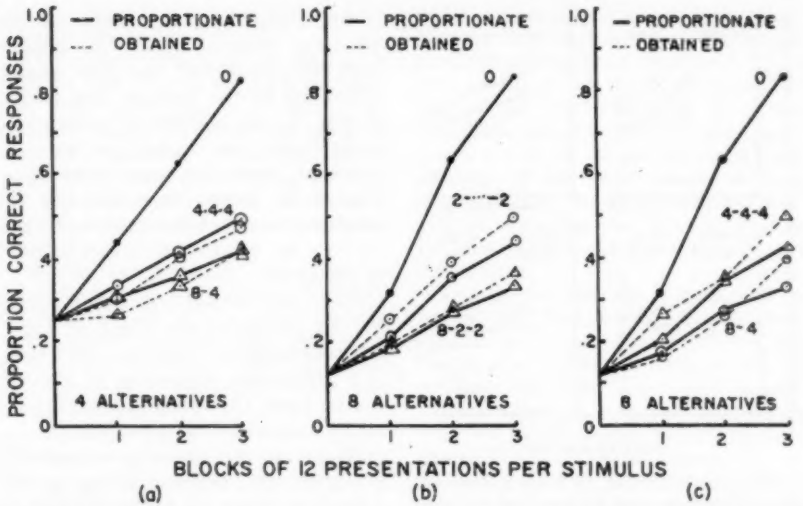

FIG. 6. Proportionate and observed response to the reduction in feedback information for the first three blocks of trials for those conditions of Exp. III for which the differences between the proportionate and observed were least. (Part a contains four-alternative conditions, while Parts b and c contain eight-alternative conditions.)

9. (12, 6, 3, 3, 0, 0, 0, 0), $I = 1.12$
10. (12, 6, 1, 1, 1, 1, 1, 1), $I = 1.20$
11. (12, 8, 4, 0), $I = .58$
12. (12, 8, 4, 0, 0, 0, 0, 0), $I = .87$
13. (12, 8, 2, 2), $I = .59$
14. (12, 8, 2, 2, 0, 0, 0, 0), $I = .88$
15. (12, 8, 1, 1, 1, 1, 0, 0), $I = .90$

To refer to conditions briefly, a shortened notation listing only noise lights will be used, e.g., 4-4-4 or 8-1-1-1-1. Conditions with nearly rectangular distributions of noise lights such as (12, 2, 2, 2, 2, 2, 2, 0) will be abbreviated further, e.g., 2- . . . -2. A "0" will be used to refer to condition with no noise.

Subjects and stimuli.—Ten *S*s served in each condition. One of these *S*s was a replacement for an *S* who apparently failed to understand the instructions. The same criteria for rejection were used as in Exp. I. The eight stimuli were the same as those used in Exp. II.

Results

Initial rate of learning.—In Fig. 6 are presented the curves for those six noise conditions which had the smallest normalized deviation areas and the two noise free conditions. Part a gives results for conditions with four response alternatives; Parts b and c for those with eight response alternatives. The normalized deviation areas for these conditions are given in Table 3. The average percentage error in the prediction of proportionality for these six conditions is 10.6. It can be observed that the normalized deviation areas were smallest when the noise was either distributed rectangularly or when the noise was heavily concentrated on a single position.

TABLE 3

NORMALIZED DEVIATION AREAS FOR EXP. III

Noise Distributions	4-Alternative Conditions	8-Alternative Conditions
2-2-2-2-2-2	—	2.45[a]
4-4-4	1.39[a]	2.50[a]
6-1-1-1-1-1-1	—	3.33
6-3-3	3.24	6.33
6-6	4.42	3.52
8-1-1-1-1	—	2.87
8-2-2	5.51	1.70[a]
8-4	2.34[a]	1.45[a]

[a] These conditions are presented in Fig. 6.

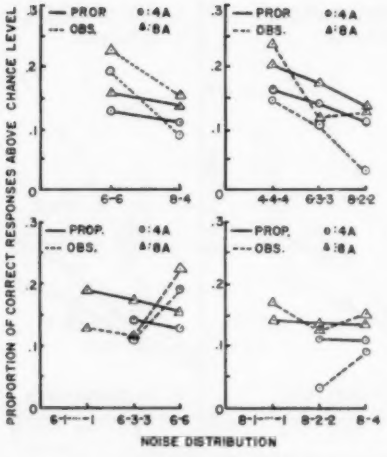

FIG. 7. Proportion of correct responses above chance level as predicted in the basis of proportionality and as observed for the conditions of Exp. III for the average of the first three blocks of trials. (Four-alternative conditions are symbolized by having data points included in circles while triangles are used for eight-alternative conditions.)

In Fig. 7 are presented the average predicted and average observed proportion of correct responses above chance level for the first three blocks of trials. To obtain these graphs, the proportion of correct responses which could be expected if *S* were responding randomly (chance level) was subtracted from the proportion of correct responses for each of the first three blocks of trials. This gave the proportion of correct responses above chance level for each block of trials. These results were then averaged over the first three blocks of trials. The points on the abscissa in Fig. 7 represent conditions. The conditions were arranged so that the predicted proportionate response values would always have a negative slope.

In the upper two graphs of Fig. 7 it can be seen that when the number of positions a noise lamp can occupy is kept constant, the slope of the observed results agrees in general with the predicted negative slopes. When, as in the bottom two graphs in Fig. 7, the number of positions a noise lamp can occupy is not kept constant, there are some marked reversals in slope, especially in the condition in which one incorrect alternative is reinforced exactly 50% of the time.

The normalized deviation areas for the points on the graph in Fig. 7 are presented in Table 3. The average percentage deviation of the obtained from the predicted results for all 13 noise conditions is 15.9.

Shape of the learning curve.—No graph for the MaxΔR values is presented because of the difficulty of presenting 15 conditions clearly. The results concerning the location of MaxΔR in the learning sequence are essentially the same as those in Exp. I and II:

$$P_0 < \frac{1}{6^8} < .001$$

$$P_1 < \frac{1}{5^5} < .001$$

$$P_{Diag.} < \frac{1}{6^{11}} < .001$$

The tau correlation between *I* values and values of MaxΔR was .543.

DISCUSSION

The most important conclusion to be drawn from these experiments is that it is possible to use a measure of feedback information *not* based on performance which permits fairly accurate predictions of initial performance on a learning task from a knowledge of only (*a*) performance in the task with no reduction in feedback information and (*b*) the amount by which the feedback information is reduced. Such predictions are not always very good, and indeed at times may be quite poor, but on the average they offer an

initially acceptable basis for making quantitative predictions.

The second important conclusion is that the value of the feedback information affects that point at which performance shows its maximum increase in correct responses. Thus, there clearly seems to be some relationship between the value of the feedback information and the shape of the averaged learning curve.

While the preceding are the two principle conclusions of this set of experiments, there are several other findings which deserve mention. The first is more of a qualification than a finding. Since predictions of initial performance were independent of number of alternatives, the important part of the calculation in the prediction rested upon the expected number of trials required to identify the correct alternative for a given stimulus. Thus, the information measures used in making predictions for a fixed number of alternatives were directly dependent upon the expected number of trials over which information had to be retained.

Secondly the proportionality prediction lead to different errors in the case of four-alternative and eight-alternative conditions. In almost all cases, predicting a proportionate reduction for the eight-alternative conditions resulted in underestimating the obtained results. In many of the four-alternative conditions, a decrease in performance proportionate to the decrease in information fed back following a response would result in an overestimate of the obtained results. Thus, it is possible that the proportionate predictions might have been best for conditions employing approximately six alternative responses. Certainly, in the further developments of any model, the number of alternative responses must be considered as an important variable affecting the performance results of the reduced information.

With regard to these differences between four- and eight-alternative conditions it is also interesting to note the relative accuracy of the predictions based upon two different types of information measures in Exp. II. There were two conditions, Q and H, for which the I_1 and I_2 values gave different predictions. For the four-alternative conditions, I_2 gave better predictions than the I_1 values and conversely for the eight-alternative conditions, (Table 2, Fig. 3). The predictions based on I_2 values essentially placed stress on changes in noise lamp position. If the noise is more confusing when there are fewer alternatives available, it is possible then that changes in the noise lamp position assume a greater significance.

The greater difficulty experienced by *S*s in the four-alternative conditions cannot be related entirely to the fact that in some conditions there were fewer positions the noise lamp could assume. Were this true, it could have been argued that the noise lamps created greater generalization between alternatives by appearing in the same position more often. In Exp. III the number of positions the noise lamp could assume and the number of times it assumed each of these positions were controlled so that any generalization which occurred between lamps should have been equal for both the four- and eight-alternative conditions. Yet, as was observed in Table 3, even under these conditions, *S*s in the four-alternative conditions found the learning to be more difficult. There are two possible explanations. The first is that the simple availability of more positions the noise lamp could assume in the eight-alternative conditions, even though it never assumed more positions than in the four-alternative conditions, made their discrimination simpler. The other is that the need for *S* to associate *two* switches with each stimulus in the four-alternative conditions made the noise more effective in interfering with learning in these conditions.

There were also, on the other hand, certain similarities between the effects of the noise on performance for both four- and eight-alternative conditions. The most noticeable one is the U-shaped function which appears in Table 3. A prediction of proportionality was best when the distribution of the place of occurrence of the noise was either very close to a rectangular distribution, or highly concentrated in a single position.

The differences found between Cond. Random$_1$ and Random$_2$ in Exp. II may be partly accounted for in this way. Such a U-shaped function seems reasonable because an even distribution of the place of occurrence of a noise lamp makes the constant lamp more discriminable, while a concentration on a single position increases the discriminability of that single incorrect position for the remainder of the incorrect positions. This is in accord with Morin's (1955) finding that although awareness of misinformative feedback increased persistence in specific incorrect modes of response it also increased learning rate.

In these experiments again, a reinforcement of 50% produced some deviations from what could be described as the "overall pattern." In Exp. III, the largest deviations from proportionality occurred when one incorrect response was reinforced exactly 50% of the time. When only one alternative was reinforced exactly 50% of the time, there was always, for both the four- and eight-alternative conditions, less than proportionate use made of the feedback information. Reinforcing an incorrect alternative 50% of the time apparently attracted more responses away from the constant lamp than did, for example, a reinforcement of an incorrect alternative 66.7% of the time. (This may account in part for the previously described U-shaped function.) A similar result occurred in a study of the effects of increments of reinforcement (Shelly, 1960). There it was shown that the greatest decrease in responding to an alternative which was reinforced 75% of the time occurred when the probability of reinforcement on another alternative reached 50%.

Summary

Three experiments are reported, each aimed at providing a partial answer to the following two questions: (*a*) When the information fed back to *S* concerning which of four (or eight) alternatives is correct, is reduced, is the initial rate of learning reduced proportionately? (*b*) Will the point at which maximum increase in proportion of correct responses takes place vary with the amount of information available to *S*, i.e., will the "shape" of the learning curve vary with amount of information fed back?

Three experiments, based upon a paired-associate paradigm, were conducted. Experiment I varied the amount of information, or conversely the amount of noise, in the feedback given *S* through variations in the number of "noisy events" which could follow a response; Exp II varied the sequential pattern of the noise; and Exp. III varied the distribution of the noisy events over the set of available incorrect alternatives.

The conclusions were that a prediction that noise reduces initial learning rate proportionately to the amount of noise can serve as a first approximation, and that there are differences between conditions in the point at which the maximum increase in proportion of correct responses occurs. The prediction of proportionality in initial learning rate was most accurate when there was a single noise lamp which was distributed evenly over the set of incorrect alternative positions, or when the noise lamp appeared very often in a single position. Consistent differences were noted between the effects of the noise in four- and eight-alternative conditions. It was more effective in depressing performance in the four-alternative conditions. For the range of conditions used, the following was generally found concerning variations in the point of maximal increase in correct responses: When such points of maximal increase were compared with feedback information, the relationship was quite close to linear.

References

Bourne, L. E., & Pendleton, R. B. Concept identification as a function of completeness and probability of information feedback. *J. exp. Psychol.*, 1958, **56**, 413–420.

Morin, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49**, 343–351.

Neimark, E. D. Effects of type of reinforcement and number of alternatives in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–228.

Shelly, M. W. Patterns of decrease in response probabilities. *J. exp. Psychol.*, 1960, **59**, 345–350.

Trowbridge, M. H., & Cason, H. An experimental study of Thorndike's theory of learning. *J. gen. Psychol.*, 1932, **7**, 245–258.

(Received April 14, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 223–233

BALL THROWING RESPONSES TO PHOTOGRAPHICALLY PORTRAYED TARGETS[1]

PATRICIA CAIN SMITH AND OLIN W. SMITH

Cornell University

This paper reports four experiments which were performed to determine how well photographs would serve as the distal stimuli for motor responses, e.g., tossing of balls to absolute distances as presented monocularly in the space immediately adjacent to *S*. The experiments are intended to provide an inductive base for an important but neglected class of perceptual problems: The determination of conditions of observation and visual stimuli which are conjointly necessary and/or sufficient for perceptions of absolute distances in the everyday world of *S*. Present theories (Gibson, 1950; Hebb, 1949; Ittelson, 1961; Koffka, 1935; Ogle, 1950) simply ignore this class of problems and in so doing they ignore problems of direct relevance to the everyday life of *S*. This unfortunate fact can be attributed to three causes.

The first is the general failure of theorists to regard *S*'s motor functions as determinants of many of the units by which *S* perceives nearby absolute distances in his everyday life. The assumption which we make here (for which there is abundant evidence) is that the space which is of greatest importance to the mobile *S* on a ground surface is the space of his immediate environment. It is in this space that *S* walks, jumps, and finds his daily bread. It is in this space that the most common instances of perceptions of absolute distances take place. Most of these perceptions involve motor acts. For instance, a puddle is perceived as too large to step across; a ditch is perceived as too broad to jump; a vacant bit of curbing is seen as too short in which to park a car.

The units of perceived absolute distance vary according to the motor acts by which *S* copes with his immediate spatial environment. The units are standard only in nomenclature. They are highly specific to each *S*; they vary within *S*s over time; they differ from *S* to *S* at any given moment. Yet, these are the units of perceived absolute distance which are most practiced in everyday life and most necessary for self-preservation. The units do not lend themselves readily to study, particularly under conditions permitting *E* the option of controlling the conditions of observation as well as selecting and varying the visual stimuli for responses. This difficulty alone may be the sufficient explanation for the general neglect of theory and research in this area.

A second cause for the failure of theory to account for absolute judgments is that light to the eye has been most frequently taken as the independent variable of visual perception (see Zajac, 1959, who would restrict visual perception to the study of perceptions which can be accounted for by ocular functions alone). Let us emphasize here that all other things equal, *S*'s posture and the position of his eye in space over time control the geometrical perspectivities of light which can appear on the retina in one-to-one correspondence with the distal stimuli of everyday life. The light entering the eye of *S* is dependent,

[1] This research was supported in part by Contract AF18 (600-443) between Cornell University and Human Resources Research Laboratory, Bolling Air Force Base, Washington, D. C., and Contract Nonr 401(14) between Cornell University and the Office of Naval Research. The assistance of P. Herman and A. Steinschneider is gratefully acknowledged.

then, upon the independent variables of *S*'s posture and motion in space (assuming that the *S*'s head is rarely static in everyday life). The *S* that is supported by a ground surface is usually aware of his absolute changes in spatial position independently of vision or concomitantly with the visual stimuli which confirm this change in position. Visual perception as studied traditionally, however, takes the visual stimuli as the independent variables. In order for the theory of visual space perception to be optimally relevant to the perceived absolute distances which are crucial for everyday life, theory must take into account the joint changes in *S*'s position in space along with the dependent changes in his visual stimuli.

A third cause has been the absence of any suitable apparatus permitting: (*a*) the selection, control, and variation of visual stimuli in conjunction with the postures which are customary (or nearly so) for the distances to be viewed; (*b*) postures appropriate to these distances. For example, the apparatus for exposing the visual stimuli for the space near the feet of *S* should require an appropriate tilt of *S*'s head for the purpose of aiming his eye at something that he sees.

The apparatus for satisfying these conditions for binocular vision in the mobile *S* might be difficult to achieve. A simple and inexpensive apparatus for monocular and static views, however, seemed possible. The research on photographically portrayed spaces suggested the use of suitably mounted photographs as the distal stimuli for motor responses corresponding to perceived absolute distances in the portrayed spaces. Monocularly viewed photographs would also permit (*a*) the study of visual direction as a stimulus for perceived absolute distances, and (*b*) the assessment of perceptions of absolute distances that are independent of the cues of accommodation and convergence. These latter cues could only carry information about the optical distances of the photographs, not of the portrayed distances; any effects attributable to accommodation and convergence would be interference effects. A test of the effectiveness of such an apparatus for motor responses to distances perceived under "optimal" visual stimulation in photographs would serve as a criterion for evaluating the extent to which other combinations of stimuli would be sufficient for perceived absolute distances.

Schlosberg (1941) demonstrated that photographs can induce compelling and highly realistic impressions of depth when *S*'s monocular field of view is restricted to a photograph. The heightened realism of this effect is probably due to the fact that when *S*'s field of monocular view is not restricted to the photograph, *S* perceives two spaces, the space portrayed in the photograph and the three-dimensional space immediately adjacent to *S* of which the photograph as an object is only a part. Under the normal conditions in which photographs are viewed, the two spaces, that is, the portrayed space and the space of *S*, are incompatible. Restriction of the field removes this incompatibility.

Photographs can also subtend to the eye visual angles which closely approximate those subtended by the objects themselves to the lens of the camera (Henney & Dudley, 1951, pp. 104–106). Gibson (1954) assumed that if photographs were viewed appropriately, the resultant perceptions of distance should be equivalent to those induced by the original scenes. This assumption was tested and found tenable (but subject to constant errors) by Smith and Gruber (1958) who compared directly the distances perceived in a still photograph with those perceived in the original space. We therefore assumed that when other factors were held constant, photographs of targets in the space immediately adjacent to *S* could be substituted for direct views of the targets themselves.

The distance to which a ball is tossed at a photographically portrayed target is an excellent distance response with which to undertake the investigation of this neglected area since (*a*) the extent of the toss is an absolute measure of the distance response, (*b*) most male adult *S*s are well practiced in making this

response to nearby targets in their immediate environment, (*c*) the response can be made to the different distances of *S*'s immediate environment, and (*d*) the response can be made under restricted as well as unrestricted conditions of viewing.

The use of photographs, however, as the distal stimuli for targets of tossed balls involves variables in addition to those of perceived distances. For instance, the degree of restriction of *S*'s monocular field of view is important in that a sufficiently large field of view can induce the impression of a space which is immediately adjacent to *S*'s person (Smith, 1958). Conversely, a minute field of view can create the impression that the viewed scene is detached and not a part of *S*'s immediate environment. Consequently, three different sizes of field of view were studied.

Another variable is the visual direction of a target. Bailey (1954) demonstrated that the visual direction of an elevated target can be given in perception by a single point of light in an otherwise dark environment. Hence, the traditional cues for depth and distance of a target relative to *S* are not necessary for establishing its direction. A different case is that of the visual direction of a target resting on a ground surface that is the basis of *S*'s support. In this instance, the absolute distance of the target from *S* is also a determinant of the visual direction of the target. Photographs displayed so as to cause *S* to tilt his head also establish for him the visual direction of the target as portrayed on a ground surface common to *S* and the target. Are the conditions of establishing the visual direction of targets in photographs sufficient for the perception of the absolute distances of the targets? This question was therefore investigated both as a control for the picture experiment and as a problem in its own right.

Experiment I

This experiment investigates the effects of introduction of three sizes of apertures in a monocular field of view on the accuracy of tosses of balls at directly viewed targets.

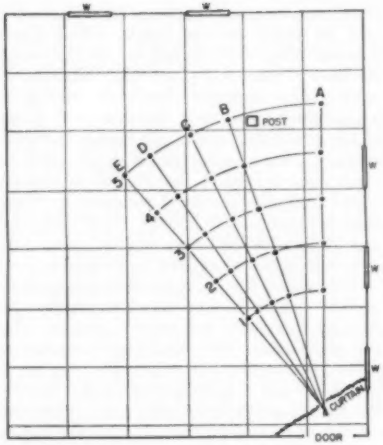

FIG. 1. Diagram of room of the experiments. (Dots indicate the target positions.)

Method

The experimental room was 36 × 30 ft., and illuminated by fluorescent tubes in industrial fixtures. The floor was made of concrete blocks 5 ft. square and was dusted lightly to make clearly visible the point of impact of the tossed balls. There were 25 target locations at five distances: 3, 4.25, 5.5, 6.75, and 8 meters from a point directly under the eye of the viewing apparatus. These will be called Distances 1 through 5, respectively (see Fig. 1). For each of the distances, five stations were located at different angles from the viewing position. A flat yellow target, 8 in. in diameter, was placed on the station specified for each trial.

When *S* entered the room (see Fig. 1) he saw only a curtain of black cloth and an eye piece which extended through a slit in the cloth. Behind the cloth was a rectangular wooden stand which was 72 in. long, 45 in. high, and 12 in. broad. The stand was located approximately parallel to and in the middle of the distance allocated to the curtain in Fig. 1. In the middle of the top plank of the stand was mounted the head of a tripod. The head of this tripod could be swivelled in either or both the horizontal and the vertical and locked into position. To this

head was attached a rectangular funnel which was jet black on the inside. This funnel tapered from 2 in. × 1¾ in. at the mount for the eye piece to a plane which was 18.5 in. wide × 15.5 in. high. From *S*'s side of the apparatus he saw only the first inch or two of the protruding end of the viewer through a curtain. Rectangular slides determined the three sizes of field of view, which in terms of visual angles for the small, medium, and large apertures were 7°12′ × 5°12′, 24°7′ × 17°54′, and 41°7′ × 30°38′, respectively. The slides in a slide holder were 20.8 in. from the eye.

Below and to the right of the funnel the curtain could be opened and closed, and through this opening *S* made his tosses. For the monocular and binocular unrestricted series the rectangular stand was moved 20 in. to the left. The curtain to the right of the funnel could be opened and closed and through this opening *S* could view the targets with unrestricted head motion for controlled periods of time. As before, he tossed his balls through a controlled opening on the lower side. A regulation soft ball weighing 6½ oz. was used. The target was centered in the field of the viewing apparatus for all aperture trials. The *S*s did not observe this process.

The *S*s for Exp. I were 50 inductees from Sampson AFB, New York. Each was given practice and instruction in tossing a ball at a target underhanded, in a space different from that of the experiment. He was made familiar with positioning his eye at the eye piece, and was given practice in tossing as hard as he could, as well as gently, through the opening in the stand without viewing a target for his tosses while his eye was in position at the eye piece. Some practice in tossing was necessary so as to insure that *S* would release the ball before the upsweep of his hand reached the top plank of the stand on which the tripod was mounted. He was practiced in closing his eyes when the ball left his hand. To insure the absence of visual knowledge of results of his throws, *E* barred *S*'s field of view with a flag as soon as *S* tossed a ball. No verbal information was ever given *S* regarding the accuracy of his tosses. A count-down system was used to insure nearly equal times of exposure of all targets for all views through the viewing apparatus as well as during monocular and binocular unrestricted views. The *S* was instructed to toss at the count of four. Some head motion parallax was possible under all conditions during the short period in which the targets were viewed.

All *S*s made one toss at one of the five target locations at each of the five distances for small, medium, and large apertures as well as for monocular and binocular unrestricted views in that order. The order of distance and location of target for any one viewing condition was randomized for each *S*. The only restriction was that all five distances were presented under each viewing condition. One *E* instructed *S*, adjusted the viewer, and operated the flag while two *E*s retrieved the balls, measured the distances tossed, recorded, and positioned targets. At no time did *S* see the *E*s beyond the viewing stand, or the room, between throws during the restricted series.

The zero point from which measurements of distance originated was a point directly beneath the eyepiece of the funnel. When the stand was displaced 20 in. to the left for the unrestricted series, the origin of measurement was not displaced. Lateral errors of tosses were not measured since in preliminary trials these were a function of the distance of the targets and exhibited a narrow range of dispersion when compared with the dispersion of the tosses in distance.

TABLE 1

ANALYSIS OF VARIANCE, EXP. I

Source of Variation	*df*	*MS*	*F*
Condition of viewing (C)	4	916.37	2.77
Distance (D)	4	11,300.97	72.48
Ss	49	2,499.38	55.86
C × *Ss*	196	331.29	7.40*
D × *Ss*	196	155.92	3.48*
C × D	16	169.79	3.79*
C × D × *Ss*	784	44.74	
Total	1249		

* $P < .01$.

Results

All *S*s were able to toss the ball at the targets. Under aperture conditions 23 *S*s hit one or more targets for a total of 37 hits. A number of *S*s inquired about the (nonexistent) lens system or glass in the viewing apparatus.

The distance of the point of impact of the ball on the floor was converted to a percentage of the target distance. These numbers were subjected to an analysis of variance which is sum-

TABLE 2

MEANS AND *SD*s OF PERCENTAGES OF TARGET DISTANCE THROWN FOR VIEWING CONDITIONS OF THE FOUR EXPERIMENTS

Exp.	Conditions of Viewing Apertures								Unrestricted Conditions			
	Small		Medium		Large		Combined		Monocular		Binocular	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
I	100.1	12.8	99.5	15.1	99.8	13.1	99.8	12.0	96.1	10.1	96.6	9.0
II	94.4	9.7	96.3	9.2	102.4	10.7	97.7	9.1	92.4	8.6	91.7	6.7
III	103.0	23.3	104.2	25.4	106.1	26.8	104.4	22.9	101.1	9.7	97.3	7.4
IV	87.2	21.4	86.7	27.8	85.5	24.8	86.4	23.4				

marized in Table 1. The second order interactions were all significant at the 1% level. The means of all distances for the small, medium, and large aperture conditions are nearly the same and are larger than those for the two unrestricted conditions (see Table 2). According to a method of analysis suggested by Tukey (1949) for the determination of sources of variation when certain treatments are known to be different as a class from certain other treatments, most of the variation associated with conditions of viewing is attributable to the difference between aperture and non-aperture conditions (see Table 3).

Target distance was a major source of variance despite the conversion of raw scores into percentage of target distance. In order to determine which of the means were significantly different from the overall mean for distances, Tukey's (1949) method was used for breaking up treatments into groups where there is no a priori basis for grouping and where an analysis of variance shows that not all treatments are alike. By this analysis the mean for Distance 5 is the only one which is significantly different from the overall mean. The far target may have been consistently undershot because it was the end point of the series.

Discussion

Restricting the field of view of naive *S*s to the surround of their targets had the overall effect of causing them to throw harder (as if the targets appeared more distant) than when views of the

TABLE 3

MEANS AND *SD*s OF PERCENTAGES OF TARGET DISTANCE THROWN OVER ALL APERTURES IN THE FOUR EXPERIMENTS

Exp.	Distance of Target in Meters										Total	
	3		4.25		5.50		6.75		8.0			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
I	107.0	15.8	105.7	14.6	101.0	14.5	95.6	12.7	89.8	12.4	99.8	12.0
II	103.9	13.2	103.6	8.9	99.2	10.7	92.3	10.9	89.4	9.0	97.7	9.0
III	114.2	15.2	109.5	19.6	104.9	14.7	99.6	14.4	93.9	10.5	104.4	22.9
IV	101.9	33.4	92.9	22.6	87.8	29.7	79.0	21.6	70.7	16.4	86.4	23.4

targets were not restricted. Had *S*s been permitted to view the points of impact of their tosses and to correct for errors it seems very likely that the difference would have been eliminated.

The Conditions × Distance interaction was predicted in advance of the experiment since use of the motor response as a criterion for perceived absolute distance as well as the restriction of the fields of view introduced unavoidable interactions. For instance, when the angular extent of a restricted field of view is held constant for views of targets at different distances, the visible distances on the ground of the target areas vary with the distance of the target. The ratio of the distance viewed within a target area to the distance separating *S* and the nearest visible portion of a target area decreases and then increases as a function of the distance of the target (assuming a level ground common to *S* and target).

The motor function of positioning the head for inspecting a target should have its greatest cue value for distance for the target nearest to *S* since the greatest inclination of the head is required. Also, for any one field of view the ratio of the distance of surface viewed in the target area to the distance separating *S* and the nearest visible point of the target area is greatest at near position. And, lastly, in the absence of knowledge of errors of response, the distances to which balls are tossed should decrease with an increase in target distance beyond a neutral point. These comments only serve to indicate that all other factors do not remain constant when the distance of targets is increased.

Experiment II

This experiment was performed to determine whether the order of observations in Exp. I had a significant effect.

Method

The serial order of conditions of observation was reversed from those of Exp. I; i.e., the first condition of tossing was with unrestricted binocular views which gave *S* a complete view of the area in which the targets were spotted at different locations and distances. Instructions and practice were essentially the same for the two experiments. The *S*s were 16 inductees from Sampson Air Force Base.

Results and Discussion

Twelve of 16 *S*s made a total of 19 target hits in Exp. II under aperture conditions as compared with only 23 of 50 *S*s who scored a total of 37 hits in Exp. I. This difference in target hits might be attributed to either or both of two causes. The first is that the additional information gained from the binocular and monocular views carried over to performance under aperture conditions. The second interpretation that *S*s of this experiment were, as a group, more accurate ball tossers than those of Exp. I, is not tenable since neither the means nor variances for monocular unrestricted views of the two experi- were significantly different. The first possibility remains tenable.

The directions of the overall differences between aperture and unrestricted views are the same in the two experiments. An effect attributable to serial order between means of the small and larger apertures was significant by t test ($P < .01$). Effects of target distance are also parallel to those of Exp. I (see Tables 2 and 3).

Experiment III

This experiment repeated Exp. I with the exception that *S*s observed photographs in the viewing apparatus instead of the room itself. The purpose of this experiment was to compare the accuracy of responses to photographs with those of the criterion group in Exp. I.

Method

Photographs were made of all targets as viewed under the large aperture condition of

Exp. I. The lens of the camera, a Graphic View II, was placed at the eye position of *S* in the viewing apparatus. The film was panchromatic X; the prints were on K-2 paper; and the lens a 152 mm. Ektar. Uniform conditions were maintained in developing and printing. The prints were achromatic with the exception of the target which was tinted a yellow to match the color of the original target. This permitted easy identification of the targets and use of instructions nearly identical to those of Exp. I. There was no attempt to achieve high verisimilitude, but rather accuracy of angular relationships between print and scene, and comparability of conditions of viewing and instruction.

Each print was bound between two glass plates. When a slide was inserted in the slide holder of the viewing apparatus, the angles subtended to the eye of *S* by objects in the photograph corresponded exactly to the angles subtended by the original objects. The slides were backlighted by the ordinary room illumination. This backlighting caused the photographs to be darker than the real scene. The high contrast between the level of brightness of the picture and the black walls of the viewing apparatus compensated perceptually for the lower brightness level of the pictures. The viewing apparatus was adjusted for direction and elevation so that it was centered on each target before the picture of that target was inserted. The *Ss* did not witness this positioning of the viewing apparatus.

The *Ss* were 30 inductees from Sampson Air Force Base. The instructions for this experiment are those of Exp. I, with the exception that after each *S* had completed all of his tosses he was shown the slides in their rack and was asked if he recognized them. He was also asked if there was anything unusual about what he had seen through the viewing apparatus.

Results and Discussion

No *S* reported that he had seen photographs in the viewing apparatus, or noted any unusual quality of what he had seen, even its achromatic nature. In this respect, the photographs were unexpectedly successful as surrogates for the room and its targets. All *Ss* recognized that the slides were pictures of the room, without understanding the mode of their use.

Their accuracy was also surprisingly good. Thirteen *Ss* made a total of 14 target hits while viewing photographs. The variance of the combined aperture responses to pictures was significantly greater ($P < .01$) than the corresponding variance of Exp. I. This difference in variance could account for the proportionately fewer hits made by the picture group than by the groups of Exp. I and II.

The mean for each of the three apertures in which the pictures were viewed is higher than the corresponding mean of Exp. I. The mean and variance of the *monocular* unrestricted responses are assumed to be valid measures for assessing "basic" differences in ability between the groups of Exp. I and III independently of responses under aperture conditions. The variance of the two groups did not differ significantly, but their means did differ ($P < .05$). Therefore, the means of the responses to pictures should be corrected by the amount of the difference between the means of the monocular unrestricted responses before the means of the responses to pictures can be validly compared with responses to direct views of the targets under aperture conditions. The subtraction of five percentage points from the mean for responses to photographs nearly eliminates the differences between aperture means for Exp. I and III. Therefore, the mean effects of photographs did not differ significantly from aperture effects or direct views of the targets. Direct views of the targets, however, did permit some monocular motion parallax which would have accounted for the significant reduction in variance.

Motion parallax and binocular disparity were effectively eliminated as cues or stimuli for perceived depth and distance by the use of photographs.

Accommodation and convergence were made for the optical distance of the photographs. If functional, their effects would have been to decrease the magnitude of the motor responses. The means (see Tables 2 and 3) provide no evidence for such effects. Motion parallax, binocular disparity, accommodation, and convergence were eliminated as cues for the absolute distances involved in the motor responses. The *SD* for the monocular unrestricted condition was only slightly greater than for the binocular unrestricted condition. This indicates that nearly all of the increase in accuracy in the unrestricted conditions is attributable to accommodation, convergence, and motion parallax in combination, and not to binocular disparity.

The *S*s of this experiment frequently reported that they did not feel sure of the results of their tosses. This feeling of uncertainty is undoubtedly due to the absence of the above four cues.

Experiment IV

This experiment determined whether the visual direction of a target on a surface which also supported *S* would serve as a stimulus for the distance of the target from *S*. The apparatus for the aperture and photograph experiments caused *S* to tilt his head as well as to adjust its left-right position. The distance to which a ball was tossed was again used as evidence of the effectiveness of the stimuli for distance.

Method

The same viewing apparatus, aperture conditions, targets, and target stations were used as in Exp. I, II, and III. Crosshairs were centered on a slide so that the position of the intersection of the crosshairs was in the center of the rectangular field of the viewing apparatus when the slide was inserted in the slide holder. A heavy sheet of white paper was mounted so that it could either cover the exterior opening of the viewing apparatus or reveal it. With this paper cover in place *S* could see only black crosshairs against a white field. Before each judgment by *S*, *E* adjusted the viewing apparatus so that the crosshairs were centered on a target. As in the other experiments the order of distance tossed, as well as the selection of target at any one distance, was randomized for each *S*. The instructions were essentially the same as for Exp. I except that *S* was told that a straight line from his eye through the crosshairs would intercept his target on the floor. The *S* was permitted no views of the room in which were located the targets. No information was given *S* regarding the accuracy of any response. The *S*s were 10 inductees from Sampson Air Force Base.

Results and Discussion

Some *S*s had difficulty in tossing balls at the points where they estimated the targets to be. Only one target hit was scored. The means for percentages of target distance thrown decrease progressively with increases in distance of the targets. This holds for Exp. I, II, and III (see Table 3) as well, but the rate of decrease in Exp. IV was accelerated greatly. The mean for Distance 1 more closely matched those of Exp. I, II, and III. In Exp. I the aiming function of positioning the head was hypothesized to be most effective for the nearest distance since the greatest inclination of the head from the horizontal is required for inspecting a target at 3 meters. The hypothesis remains tenable in terms of the obtained mean, but is false in terms of the *SD* which is largest at the near distance. The hypothesis requires further data for evidence of its tenability.

The evidence of this experiment is that the aiming function of the head in establishing the visual direction of targets located at 3 to 8 meters is sufficiently informative to *S* (for both direction and distance) to account for a substantial proportion of mean distance to which balls were tossed in Exp. I, II, and III. The range of distance responses was from 101.9% for the 3-meter target to 70.7% for the 8-meter target. An increase in the distance of the targets

would probably have caused the mean percentage distance of response to shrink much more.

The positioning of the eye for inspecting a nearby target on the ground is a normal function of *S*'s posture. This posture need not be exact since the vertical and horizontal motions of the eye that supplement postural adjustment complete the process of establishing the visual direction of the target. As the distance of ground targets passes through the space immediately adjacent to *S* the necessity for changes in the tilt of the head diminishes since relatively small vertical rotations of the eye can then scan, without strain to *S*, targets located as far away as the horizon. Theoretically, the informative value of visual direction as a stimulus for distance would rapidly become zero as the line of regard approached the horizontal.

All of the *SD*s for this experiment were greater than the corresponding *SD*s for target distances in Exp. I, II, and III (see Table 3). Even though the responses to distances of the targets in this experiment were not due to chance, the *SD*s were large enough so that any statement about visual direction as a stimulus sufficient for everyday survival behaviors of *S* would be exceedingly optimistic.

The visual perceptions of distance required for motor responses to targets in the immediately adjacent space of *S* is clearly a complex process. Traditional studies of the perception of distance take into account only the depth cues or stimuli found in the light as projected on the retina. The principal finding of this experiment is that when all other factors are constant, the function of aiming the eye by causing the head to be positioned is a determinant of perceived distance (assuming a level ground common to both *S* and the nearby targets).

Discussion

The effect of target distance was common to the four experiments. The *S*s undershot the far targets and overshot the near targets (3 meters). The mean effects over apertures of increasing target distance was a decrease in the mean percentage of target distance of the tossed balls. There were no reversals of the trend, which held also for the monocular and binocular unrestricted distance responses. The distances of the targets were all well within the physical capabilities of tossing of all *S*s, as demonstrated in their familiarization trials with the apparatus. The most probable cause for the tendency to undershoot far targets was the absence of visual or verbal knowledge of results for a task which would probably improve in accuracy with formal training.

Accommodation and convergence in the photographic series of Exp. III were made for the distance at which the photographs were presented while the concomitant perceptions were of the spaces and distances for which the photographs were surrogates. Accommodation and convergence were necessary for the perceptions, but the latter were otherwise independent of these ocular functions. The spatial significance of the light dominated the perceptions. A tenable inference is that accommodation and convergence are probably of no effect as cues for the perceived depths and absolute distances over surfaces which are characteristic of everyday life. This does not mean that accommodation and convergence cannot be demonstrated to be sufficient for perceptions of depth relationships of small objects or points. The relevance of such demonstrations to the spatial perceptions of everyday life remains to be empirically determined.

The only theory to which these data may be relevant is the S-R theory of Gibson (1950, 1954) who would account for perceptions of space and distance in terms of gradients of light in the optic array. Given the gradients of light as transmitted by the photographs in Exp. III, the Gibson theory would predict "veridical" perceptions of depth and distance. The data of these experiments indicate that Gibson's theory is both adequate and inadequate depending upon the criteria which are used to

evaluate it. Perceptions of both depth and absolute distances were obtained in Exp. III, thus confirming Gibson's theory. On the other hand, the most important question is whether the theory is adequate to account for the visual perceptions of absolute distances in *S*'s immediate environment. At this point the answer must be no, since the *SD*s for the distance responses to the portrayed targets are nearly twice those of the aperture conditions in Exp. I. It is, of course, possible that practice with verbal correction under the conditions of Exp. III and IV would cause the *SD*s of the distance responses to equal or be less than those of the aperture conditions of Exp. I. Granting this possibility it should be evident that the ultimate theory should be one which will account for the "nature" that *E* samples as well as the "nature" that he can create.

The data of Exp. IV demonstrate that distance responses dependent upon visual direction can be made in the absence of gradients of light. For these data the Gibson theory is irrelevant and therefore not necessary. The introduction of the light from the photographs, however, increased the number of target hits as well as the mean percentages of target distances of the tosses. Certainly the visual world adds to the kinesthetic world information important to the ability to perform.

The changes in perspectivities of visual stimulation which are dependent upon the absolute changes in spatial position of *S*'s eye were immobilized by the fact that head movement had no such effect with the photographs of Exp. III. The static images of the photographs served to confirm the posture taken by the head in positioning the eye, but not the minor changes in position of the eye during observation. This degraded the visual stimuli which are usually available in everyday life even though the light to the eye was in one-to-one or nearly one-to-one correspondence with the distances of the targets on the floor. The decrease in accuracy of the distance responses resulted.

If changes in the light to the eye confirm changes in the absolute spatial position of *S*'s eye, how does the light acquire significance for absolute distance responses such as ball tosses? The visual knowledge of the results of a tossed ball involves perceptions in addition to those of absolute distance. For instance, the perceived location of the point of impact of a ball with respect to a target is not dependent upon the perceived absolute distance separating the point of impact and target, e.g., the impact may be perceived as "under" or "over." The perception of "under" may result in additional force being applied to the succeeding toss. Should the perceived location of the second toss satisfy *S*, the perceived force of that toss would then become the unit of perceiving the absolute distance of the target. According to this rationale, appropriate units of perception for absolute distances can evolve as a consequence of perceptions of change of relative location of objects brought about by changes in motor functions. Hence, changes in the perspectivities of light on the retina, other things equal, may confirm not only changes in the spatial location of *S*'s eye, but also changes in the relative location of objects as brought about by other functions of *S*'s body.

Summary

The evidence from four experiments involving ball tossing demonstrates that: (*a*) Photographs, when viewed appropriately, can serve as the distal stimuli for the motor response of tossing a ball to the distances of targets in the space immediately adjacent to *S*. (*b*) Restriction of the field of view of the viewing apparatus increases the extent of the perceived distances as defined by the motor responses. (*c*) Accommodation, convergence, motion parallax, and binocular disparity in combination do not have to be in correspondence with the distances to which accurate motor responses (target hits) are made. (*d*) Visual direction alone is an insufficient basis for accurate motor responses to the distances of the targets. (*e*) Photographs may be a useful tool for controlling the light to the eye for the study of visual

stimuli for perceptions of absolute distances relative to *S*.

REFERENCES

BAILEY, N. J. Determination of the location of the center, or centers of projection in anomalous correspondence. Unpublished doctoral dissertation, Ohio State University, 1954.

GIBSON, J. J. *Perception of the visual world.* Boston: Houghton Mifflin, 1950.

GIBSON, J. J. A theory of pictorial perception. *Audiovis. commun. Rev.*, 1954, **2**, 3–23.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HENNEY, K., & DUDLEY, B. *Handbook of photography.* New York: McGraw-Hill, 1951.

ITTLESON, W. H. *Visual space perception.* New York: Springer, 1961.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

OGLE, K. N. *Binocular vision.* Philadelphia: Saunders, 1950.

SCHLOSBERG, H. Stereoscopic depth from single pictures. *Amer. J. Psychol.*, 1941, **54**, 601–605.

SMITH, O. W. Comparison of apparent depth in a photograph viewed from two distances. *Percept. mot. Skills*, 1958, **8**, 79–81.

SMITH, O. W., & GRUBER, H. Perception of depth in photographs. *Percept. mot. Skills*, 1958, **4**, 307–313.

TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.

ZAJAC, J. L. Depth perception of stereoscopic images resulting from fusion of crossed and uncrossed images. *Amer. J. Psychol.*, 1959, **72**, 163–183.

(Received May 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 234–242

EFFECTS OF AMOUNT AND PERCENTAGE OF REINFORCEMENT AND NUMBER OF ACQUISITION TRIALS ON CONDITIONING AND EXTINCTION [1]

ALLAN R. WAGNER [2]

State University of Iowa

The relative performance levels obtained under partial and continuous instrumental reward training have been characterized by Goodrich (1959), Haggard (1956), and Spence (1960, Ch. 6) as including: (*a*) during early acquisition trials, a decrement under partial reinforcement which is more pronounced or more persistent as performance is measured closer to the goal region; (*b*) during asymptotic acquisition trials, superiority of partial reinforcement on response measures relatively early in the response chain; and (*c*) superior extinction performance following partial reinforcement.

Of primary interest in the present investigation was a determination of the way in which magnitude of reward influences the two mentioned acquisition effects, and the way in which magnitude of reward and number of acquisition trials jointly influence the extinction effect. Such determinations were thought to be particularly relevant to those theoretical interpretations (e.g., Amsel, 1958; Beier, 1958; Logan, 1960; Spence, 1960, Ch. 6) of partial reinforcement phenomena which have stressed the importance of anticipatory reward ($r_g - s_g$) and anticipatory frustration ($r_f - s_f$), both of which are assumed to vary with reward magnitude and number of acquisition trials.

[1] This research was conducted during the the author's tenure as a National Science Foundation Predoctoral Fellow at the State University of Iowa, and comprises part of a dissertation submitted to the Graduate College of that University, in partial fulfillment of the requirements of the PhD degree. The author is indebted to Kenneth W. Spence for his advice and assistance throughout the course of the investigation.

[2] Now at Yale University.

METHOD

Subjects

The *S*s were 64 male and 64 female experimentally naive hooded rats from the colony maintained by the Psychology Department of the State University of Iowa. Their ages ranged from 125 to 130 days at the beginning of the experiment.

Apparatus

The apparatus was a straight, enclosed pine runway consisting of a 15-in. entry box, a 12-in. startbox, a 33-in. alley, and a 15-in. goalbox. All sections of the runway were 3 in. wide with the exception of the startbox which was narrowed to 2 in. Opaque doors separated the adjacent sections but, with the exception of the start door, were closed only to prevent retracing. A clear glass ceiling covered all sections of the runway at a height of 4 in. The walls and floor of the entry and startboxes were painted a flat grey, those of the alley and goalbox a flat black.

The apparatus was located in a black, sound-deadened room. Dim, diffuse illumination was provided by the light reflected from the black walls and ceiling and originating from four shielded 150-w. fluorescent tubes placed 4 ft. above the runway. A large fan located directly beneath the runway provided a general masking noise.

The apparatus provided three time measures of locomotor performance in the runway. Starting time was measured from the opening of the start door to the interruption of an infrared beam located 6 in. distant. Running time was measured from the interruption of this first beam to the interruption of a second beam located 12 in. beyond the start door. Goalbox time was measured between the

interruptions of infrared beams located 7 in. and 1 in., respectively, from the end of the runway. The last beam passed just above the front lip of a stationary, ¾ in. deep, metal food cup.

Procedure

Experimental design.—The experimental design included two levels of three variables: percentage reinforcement (100% vs. 50%), magnitude of reward (.08 gm. vs. 1.0 gm.), and number of acquisition trials (16 vs. 60). The design may be schematized as a 2×2×2 factorial, 16 *Ss*, 8 males and 8 females, being randomly assigned to each of the eight experimental groups.

Habituation.—Thirty days prior to the first training day *Ss* were placed on a 24-hr. food deprivation schedule, with 30-min. access each day to an ample supply of wet mash prepared from ground Wayne Lab Blox. This maintenance schedule was employed throughout the course of the experiment, while water was continuously available except during the experimental sessions.

On the 2 days prior to experimental training *Ss* were allowed to explore the runway in groups of three for 5-min. periods. After the exploration on each of these days *Ss* were placed in individual grey goalboxes and allowed to eat a number of reward pellets from a food cup identical to that used in the runway.

Experimental training.—A training trial was initiated with the introduction of *S* into the entry box. When *S* had entered the narrow startbox the retrace door was lowered and 3 sec. later the start door was raised. (The 2-in. width of the startbox discouraged *S* from turning so that all *Ss* remained oriented toward the start door during the 3-sec. interval.) When *S* had entered the goalbox the retrace door was closed and *S* was removed 20 sec. later on nonreinforced trials or immediately after eating on reinforced trials.

The *Ss* were tested in squads of 8, one *S* from each of the experimental groups. Eight such squads of males were tested first on each experimental day followed by eight squads of females. The *Ss* were tested in the same sequence in every experimental session, but the order of testing was systematically varied among the squads so that equal numbers of *Ss* from each group were assigned to each of the squad positions. Following the testing of all 8 *Ss* in a squad, they were returned to their home cages, and 30 min. later *Ss* were given their daily feeding. According to this schedule *Ss* were tested under approximately 22½–23 hr. food deprivation.

Acquisition.—Half of the *Ss* received 16 and the other half 60 acquisition trials, 1 trial per day, during which the percentage reinforcement and magnitude of reward were manipulated. The sequence of reinforcement and nonreinforcement for the partial reinforcement groups was arranged as follows: + − + − − + + − − + + − + + − − − + + + − − +. It should be noted that this sequence has the following characteristics: (*a*) two reinforced and two nonreinforced trials appear in each successive block of four trials; (*b*) the first and last trials are reinforced; (*c*) never more than three reinforcements or nonreinforcements occur in a row; (*d*) when this 20 trial sequence is repeated the first order sequential probabilities are counterbalanced, i.e., reinforcement equally as often follows a reinforcement as it does a nonreinforcement.

The reward pellets were pressed from wet mash prepared immediately preceding the running of each squad of *Ss*. Two pellet-makers were used which were designed to form pellets of .08 gm. and 1.0 gm. dry weight, when the mash was prepared by mixing .5 ml. distilled water with each gram of ground Wayne Lab Blox.

Extinction.—Extinction Trial 0 was administered to each group the day following its last acquisition trial. Extinction Trials 1 to 32 were then run on consecutive days. On each extinction trial *S* remained in the nonreinforced goalbox for 20 sec.

Results

Acquisition

Mean starting, running, and goalbox speeds over blocks of four acquisition trials are plotted in Fig. 1 for the treatment groups receiving the four combinations of percentage reinforcement and magnitude of reward. Measures from all *Ss*, i.e., 16- and 60-trial groups combined, are included in the curves over the first four blocks of acquisition trials, whereas the subsequent blocks of trials represent means from only those *Ss* in the 60-trial groups. Over the first 16 trials, in all three response measures, large reward produced faster speeds than did small reward, while continuous

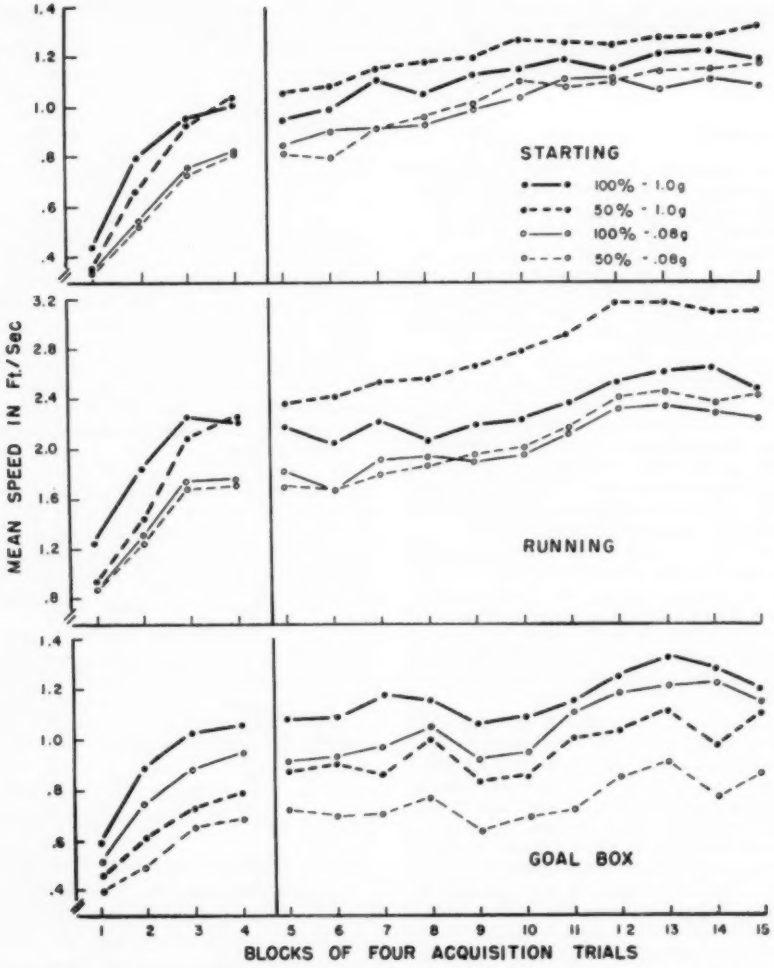

FIG. 1. Mean acquisition speeds for the treatment groups differing in magnitude of reward and percentage reinforcement. (The vertical line marks the point at which training is terminated for half the *S*s in each group.)

reinforcement produced faster speeds than did partial reinforcement.

In order to evaluate the reliability of these observations, as well as the equivalence of the 16- and 60-trial acquisition groups, a separate analysis of variance appropriate for a factorial design was computed for the means over Trials 1 to 16 for each of the three response measures. Since half of the *S*s in each of the groups were male and half female, sex was also included in these and all subsequent analyses as a "control" variable.

The results of these analyses showed that: (*a*) magnitude of reward was significant beyond the .005 level in the starting, running, and goalbox measures ($F = 26.67$, $F = 23.74$, and $F = 8.20$, respectively; $df = 1/112$); (*b*) percentage reinforcement was highly significant in the goalbox measure ($F = 48.35$, $df = 1/112$, $P < .001$), reached significance in the running measure ($F = 4.05$, $df = 1/112$, $P < .05$), but failed to approach significance in the starting measure ($F < 1$); (*c*) although the starting and running speed curves suggest an interaction between magnitude of reward and percentage reinforcement, this tendency was not statistically reliable in any of the response measures; and (*d*) there were no reliable differences between the experimental groups receiving only these 16 acquisition trials and those which subsequently received 44 additional trials.

With reference again to Fig. 1, it is apparent on all three response measures that each of the large reward groups maintained its advantage over the corresponding small reward group throughout the 60 acquisition trials. Likewise, in the case of the goalbox measure the 100% reinforcement groups maintained their advantage over the 50% groups. However, on both the starting and running measures the performance of each of the 50% groups may be seen to surpass that of the corresponding 100% group by the end of training.

In order to evaluate statistically the asymptotic performance levels attained under the various experimental treatments received by the 60-trial *S*s, a separate analysis of variance appropriate for a factorial design was computed for the means over the last 8 acquisition days on each of the response measures. The results of these analyses showed that: (*a*) the effect of magnitude of reward attained significance in both the starting and running measure ($F = 6.05$, $P < .025$, and $F = 9.53$, $P < .005$, respectively; $df = 1/56$) but is associated with a probability between .10 and .20 in the goalbox measure ($F = 2.67$, $df = 1/56$); (*b*) 50% reinforcement led to significantly higher performance levels than did 100% reinforcement on both the starting and running measure ($F = 4.03$, $P < .05$, and $F = 5.03$, $P < .025$, respectively; $df = 1/56$), whereas 50% reinforcement led to a significantly lower performance level than did 100% reinforcement on the goalbox measure ($F = 8.01$, $df = 1/56$, $P < .01$); and (*c*) in none of the three response measures was there a statistically reliable interaction of percentage reinforcement and magnitude of reward.

Extinction

Since the several experimental groups began extinction at markedly different levels, a comparison of the obtained extinction speeds appeared to offer only minimal information about the rate of extinction. A preferable index appeared to be extinction speed expressed as a proportion of the corresponding terminal acquisition speed. Consequently each response speed for each *S* over Extinction Trials 1–32 was expressed as a proportion of the mean speed *S* had attained on that response measure over the last four acquisition trials and Extinction Trial 0.

Figure 2 presents the mean starting, running, and goalbox proportions over blocks of four extinction trials for the eight experimental groups. Of major concern is the comparison of the resistance to extinction of the 50% and 100% groups following the various acquisition conditions. In all

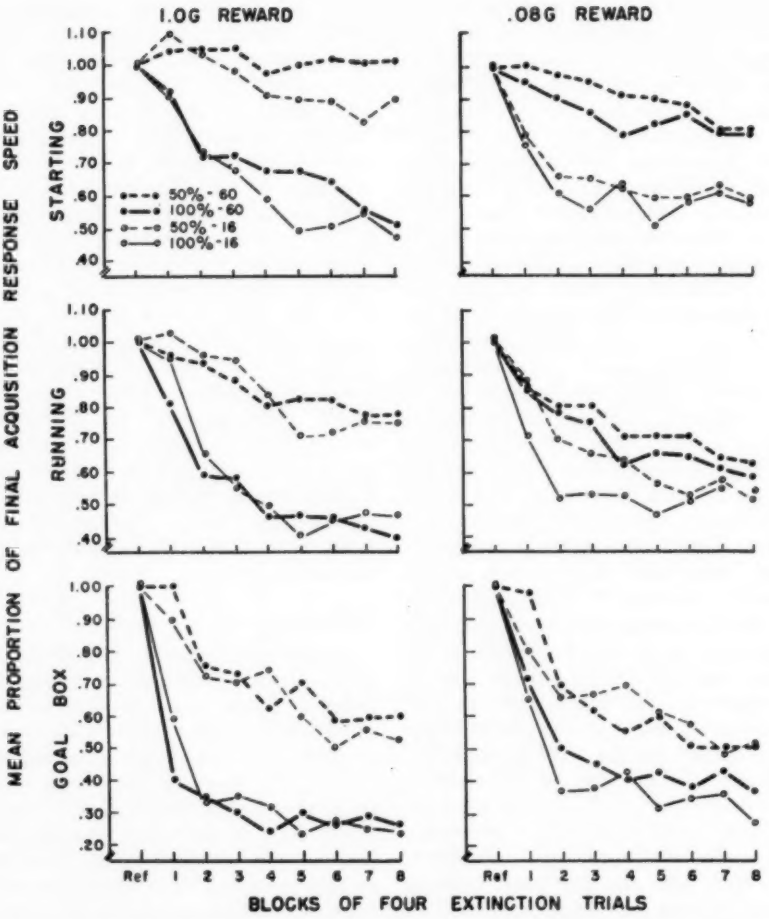

FIG. 2. Mean extinction proportions as a function of the percentage reinforced (50% or 100%) and the number (16 or 60) of prior acquisition trials plotted separately for the groups which had received 1.0-gm. or .08-gm. rewards during training. (The reference point includes the final four acquisition trials and Extinction Trial 0.)

three response measures it may be observed that with acquisition conditions of 1.0-gm. reward, 50% reinforcement led to a markedly greater resistance to extinction than did 100% reinforcement. However, when .08-gm. reward was employed, the superiority of the 50% groups, although still present in every comparison, was appreciably smaller. It should also be noted that in none of the comparisons did the difference in extinction performance between the 50% and 100% groups appear to vary with the number of prior acquisition trials.

A separate analysis of variance appropriate for a factorial design was computed for the mean proportions over Extinction Trials 1–32 for each of the response measures. The results of these analyses with respect to the percentage reinforcement variable showed that: (*a*) in none of the measures did the triple interaction of percentage reinforcement, magnitude of reward, and number of acquisition trials reach significance; (*b*) in none of the measures did the interaction of percentage reinforcement with number of acquisition trials reach significance ($F < 1$ in each case); but (*c*) in all three response measures the interaction of percentage reinforcement with magnitude of reward was statistically significant ($F = 15.70$, $P < .001$; $F = 14.97$, $P < .001$; and $F = 4.71$, $P < .05$ for the starting, running, and goalbox measures, respectively; $df = 1/112$).

Since the effects of percentage reinforcement were reliably different when combined with 1.0-gm. reward than with .08-gm. reward, separate F tests were computed of the differences between the 50% and 100% treatment groups at the two levels of reward magnitude. The results of these comparisons for the starting and running measures indicated that the 50% groups were significantly more resistant to extinction than the 100% groups when combined with 1.0-gm. reward ($F = 43.54$, and $F = 47.33$, respectively; $df = 1/112$; $P < .001$ in each case) but not when combined with .08-gm. reward ($F < 1$ and $F = 1.97$, respectively; $df = 1/112$). In the goalbox measure, however, percentage reinforcement was statistically reliable when combined with either 1.0-gm. reward ($F = 59.65$, $df = 1/112$, $P < .001$) or with .08-gm. reward ($F = 21.88$, $df = 1/112$, $P < .001$).

Also of interest are the separate effects of magnitude of reward and number of acquisition trials on resistance to extinction. As noted previously, magnitude of reward interacts with percentage reinforcement in determining resistance to extinction. Consequent to this finding, separate F tests were also computed comparing the 1.0-gm. and .08-gm. treatment groups at the two percentages of reinforcement on each of the three response measures. When combined with a 50% reinforcement schedule 1.0-gm. reward, as compared to .08-gm. reward, led to a greater resistance to extinction which attained statistical significance in the case of the starting measure ($F = 17.85$, $df = 1/112$, $P < .001$) and the running measure ($F = 10.69$, $df = 1/112$, $P < .005$), but not the goalbox measure ($F < 1$). When 100% reinforcement was employed, however, 1.0-gm. reward led to *less* resistance to extinction than did .08-gm. reward. The latter finding was significant in the case of the goalbox measure ($F = 5.71$, $df = 1/112$, $P < .025$) and the running measure ($F = 4.83$, $df = 1/112$, $P < .025$) but not in the starting measure ($F = 1.87$, $df = 1/112$, $.10 < P < .20$).

Statistical analyses revealed that an increase in the number of acquisition trials led to a reliable increase in the extinction proportions only in the case of those treatment groups which received .08-gm. reward, and then only on the starting and running measures ($F = 22.41$, $P < .001$, and $F = 5.72$, $P < .025$, respectively; $df = 1/112$) there being no reliable effect of number of acquisition trials upon the goalbox measure. Although number of acquisition trials seldom affected the extinction proportions it would still be expected to affect extinction *speeds*, since *S*s began extinction

at a higher level following 60 than following 16 acquisition trials. When analyses of variance were performed on the untransformed extinction speeds significantly superior extinction performance of those groups receiving 60 acquisition trials, as compared to those receiving only 16 acquisition trials, was obtained in all comparisons.

Discussion

Beier (1958) and Logan (1960) have assumed that under conditions of inconsistent reinforcement, $r_g - s_g$ is raised or lowered on each trial toward the value appropriate for the reward received on that trial. Several other investigators (e.g., Amsel, 1958; Spence, 1960, Ch. 6) have asserted, in addition to their individual assumptions concerning the course of $r_g - s_g$, that after some minimal number of reinforcements, nonreinforcement will elicit a primary motivational response termed *frustration*, which is directly related in intensity to the strength of $r_g - s_g$. An anticipatory or conditioned form (r_f) of this emotional response is further assumed to generalize to stimuli which antedate the goal, and is expected not only to lead to an increase in drive level, but also, through its characteristic response produced stimuli (s_f), to tend to elicit unlearned or previously learned responses which are antagonistic to the approach response. It is assumed however, that during random partial reinforcement training that these cues may become conditioned to, and tend to elicit the locomotor response itself.

While these positions when taken together lack in many instances the degree of development necessary for unequivocal deductions, they do allow for a consistent interpretation of a sizeable number of partial reinforcement phenomena. For example, during the early, prefrustration acquisition trials, partial reinforcement should lead to a lower $r_g - s_g$ than should continuous reinforcement and hence, as presently obtained, a lower level of responding on all response measures. If one makes the special assumption with Beier (1958) that $r_g - s_g$ under inconsistent reward, approaches the mean of the $r_g - s_g$s appropriate to the several rewards received, then $r_g - s_g$ under random 50% reinforcement approaches 50% of the level attained under 100% reinforcement and hence a greater partial reinforcement decrement should be obtained with large than with small reward. The present results are in this direction but do not attain significance. The fact that a similar nonsignificant tendency was observed by Hulse (1957) in the only previous attempt at such a comparison, however, suggests that it may be a reproducible effect.

The latter, postfrustration performance is less determinant, depending upon the relative degree of $r_g - s_g$ superiority of the continuous groups as compared to the $r_f - s_f$ induced drive level superiority of the partial groups, as well as upon the degree of conditioning of s_f to the approach response under partial reinforcement. Since random partial reinforcement leads commonly to greater asymptotic starting and running speeds than does continuous reinforcement, it must be argued not only that the approach response has become effectively conditioned to the early alley cues plus s_f, but also that the drive level advantage of the partial groups is greater than the $r_g - s_g$ advantage of the continuous groups.[3]

[3] That the partial superiority is not found on those response measures close to the goal may be attributed to less effective conditioning of the approach response to the goalbox cues plus s_f, and hence the persistance to some degree of competing responses elicited by s_f. This argument appears quite reasonable in the context of the prior assumptions if it is noted that $r_f - s_f$ may be expected to become conditioned to the approach response when the s_f cues are introduced initially at a weak value, and hence with a negligible tendency to elicit competing responses, and are then increased gradually at the same time that the approach response is being strengthened. This condition obtains much more clearly in the early portions of the alley than in the goal region where

Since $r_f - s_f$ is presumed to be a function of $r_g - s_g$, increasing the magnitude of reward should increase the partial reinforcement drive level superiority. The present finding of no significant asymptotic interaction between reinforcement schedule and reward magnitude might be interpreted as casting doubt on the frustration interpretation, or as lending additional plausability to the notion that the continuous reinforcement $r_g - s_g$ advantage is also increased with increased reward. However, that there was observed a sizeable and consistent tendency for the starting and running partial reinforcement asymptotic superiority to be greater with large than with small rewards appears to justify further study.

It follows most clearly from the theoretical formulation, that partially reinforced *S*s that have been trained to approach in the presence of frustration stimuli will be more resistant to extinction than will continuously reinforced *S*s that have competing responses elicited to a greater degree by the anticipatory frustration which develops during extinction. Furthermore, the present finding, as well as that of Hulse (1958), of a greater partial reinforcement extinction superiority following large than following small rewards is not unexpected. Nor is the observation in each of these studies, that this increased effect is due, in part, to *faster* extinction following continuous reinforcement with large as compared to small rewards. Since the magnitude of $r_g - s_g$ would be greater at the start of extinction with large than with small rewards, it is consistent that larger rewards should lead to a slower rate of extinction for partial *S*s due to increased motivation, but should lead to a faster rate of extinction for continuous *S*s owing to the greater vigor of the competing responses elicited by the more intense s_f.

On both theoretical (Amsel, 1958) and empirical (Amsel, 1958; Koehler, 1956; Weinstock, 1957) grounds it was expected that a greater partial reinforcement extinction superiority would be observed following 60 than following 16 acquisition trials. In order for the present failure to observe such a tendency to be considered consistent with frustration theory it must be proposed that s_f had already become effectively conditioned to running in the partially reinforced groups by the end of 16 trials. The present results would more clearly appear to place restrictions on the assumptions that may be made concerning the nature of the contributions of number of acquisition trials and magnitude of reward to the partial reinforcement extinction effect. While it may be assumed that an increase in the number of acquisition trials under certain conditions may lead to an increase in the magnitude of the partial effect, it must also be assumed that the function expressing this relationship has a limiting value determined by the magnitude of reward employed.

While it is not unlikely that both frustration and incentive motivation are important determiners of partial reinforcement phenomena, the present formulations are necessarily speculative. Further clarification awaits less equivocal evidence of the relative values of these factors in the various portions of training and at different segments of the response chain.

Summary

The acquisition and extinction of locomotor behavior in various segments of a straight alley were investigated as a function of percentage reinforcement (50% or 100%), magnitude of reward (.08 gm. or 1.0 gm.), and number of acquisition trials (16 or 60). Sixteen naive hooded rats were assigned to each of the eight groups in a $2 \times 2 \times 2$ experimental design. A 24-hr. intertrial interval was observed during acquisition and during the 33 extinction trials which followed.

The relative acquisition performance levels of partially and continuously reinforced *S*s were found to depend upon the portion of training and the segment of the response

due to the proximity of the primary frustration event, generalized $r_f - s_f$ may be expected to occur earlier and to follow a course of greater intensity, and thus produce responses which compete more effectively with the approach response.

chain that was observed. When the response measure was taken over the early segments (first and second 6 in.) of the 5-ft. alley, there was observed an initial superiority of the 100% *Ss* but an asymptotic superiority of the 50% *Ss*. When the response measure was taken over the final 6 in. preceding the goal cup the superiority of the 100% *Ss* persisted over the entire course of acquisition. Large as compared to small rewards, in addition to producing higher overall response speeds, also tended to produce a greater early trials 50% decrement and a greater asymptotic 50% superiority, although no statistical support for the latter two tendencies was obtained.

Of primary interest during extinction were the effects of size of reward and number of acquisition trials on the magnitude of the commonly obtained superiority of partially reinforced as compared to continuously reinforced *Ss*. The extinction results showed that whereas the "partial reinforcement effect" did not vary with number of prior acquisition trials, it was markedly greater following 1.0-gm. than following .08-gm. rewards. The increased partial reinforcement effect with larger rewards reflected not only greater resistance to extinction of partially reinforced *Ss* but also less resistance to extinction of continuously reinforced *Ss* with large as compared to small rewards.

The acquisition and extinction results were considered in relation to frustration and incentive motivation theory.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

BEIER, E. M. Effects of trial-to-trial variation in magnitude of reward upon an instrumental running response. Unpublished doctoral dissertation, Yale University, 1958.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

HAGGARD, D. F. Acquisition and extinction of a simple running response as a function of partial and continuous schedules of reinforcement. Unpublished doctoral dissertation, State University of Iowa, 1956.

HULSE, S. H. Schedule of reinforcement, amount of reinforcement, and duration of goal confinement as variables in conditioning and extinction. Unpublished doctoral dissertation, Brown University, 1957.

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

KOEHLER, J., JR. Acquisition and extinction of a running response as a function of the percentage of reinforcement and the number of acquisition trials. Unpublished doctoral dissertation, Tulane University, 1956.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

SPENCE, K. W. *Behavior theory and learning: Selected papers.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WEINSTOCK, M. B. A factorial study of some variables affecting resistance to extinction under partial reinforcement with spaced trials. Unpublished doctoral dissertation, Indiana University, 1957.

(Received June 11, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 243–252

STIMULUS ATTRIBUTES AND DRIVE IN PAIRED-ASSOCIATE LEARNING[1]

HERBERT LEVITT[2] AND ALBERT E. GOSS

University of Massachusetts

The paradigm for acquired learning with a single paired-associate unit is homologous with the paradigm for classical conditioning (Goss, Morgan, & Golen, 1959; Spence, 1958). Because the number of units can be varied there are many combinations and patterns of relationships among stimulus and response members which evolve and which make paired-associate learning a very useful task for investigations of interactions of task with other variables. The present study was concerned with the single and joint effects on paired-associate learning of four task variables, the similarity and association values of stimulus and of response members, and of two presumed drive variables, manifest anxiety and failure stress.

Task Variables

Similarity.—Similarity of the stimuli of paired-associate units may be in a vertical fashion, within the stimulus or response members, or laterally between stimulus and response members. Such similarity has been defined in terms of common elements (Underwood, 1953a) such as the number of similar letters among the stimuli, and in terms of common responses (Gibson, 1942) and meaning (Underwood, 1953b). Increasing the numbers of common elements among stimulus members slows learning (Feldman & Underwood, 1957; Underwood, 1953a). While such increases among response members increase overt errors, no significant effects on acquisition speed has been reported (Underwood, 1953a).

Similarities of stimulus members defined by common responses is directly related to learning speed (e.g. Gibson, 1942). However, for similarity defined by common meanings, stimulus members of intermediate similarity may be learned more rapidly than those of either high or low similarity (Underwood, 1953b). Similarity of response members, in terms of common meaning, is directly related to overt errors, but may (Feldman & Underwood, 1957) or may not (Underwood, 1953b) be related inversely to learning speed.

Association values.—In general, association values of stimulus and of response members are directly related to learning speed (Jantz & Underwood, 1958; Lifton, 1956; Mandler & Huttenlocher, 1956). Cieutat, Stockwell, and Noble's (1958) results suggest that these relationships hold across all levels of association values of stimulus and of response members, while Mandler and Campbell's (1957) finding suggests that direct relationships occur at only some levels. On the whole, association values of response members seem to have greater effects than association values of stimulus members (Cieutat et al., 1958; Hunt, 1959; Noble & McNeely, 1957; Sheffield, 1946).

Sheffield's (1946) is the most systematic attempt to explain these largely direct relationships between learning speed and association values of stimulus

[1] A more extensive account of the first of the two experiments reported here can be found in Levitt (1956). Parts of these data were presented at the 1957 meetings of the EPA in New York City. Experiment II was supported by a grant from the Research Council of the University of Massachusetts. Herbert Gewirtz, Harvey Lifton, and Ira Mintz contributed to various aspects of the two experiments.

[2] Now at Department of Psychiatry, University of Wisconsin Medical School, and Wisconsin Psychiatric Institute, University Hospitals, Madison, Wisconsin.

and of response members. Stimulus members of high association values are presumed to elicit mediating perceptual responses which are more invariant, distinctive, and compatible than the perceptual responses aroused by stimulus members of lower association values. Furthermore, stimuli produced by the perceptual responses might be more important components of stimulus compounds to which responses are to be conditioned than the external stimuli which presumably elicit both perceptual responses and the responses to be learned.

These postulated aspects of association values of stimulus and of response members of paired-associate units are independent of similarity defined by common elements. Similarity and association values of stimulus members might, therefore, combine to influence paired-associate learning. Also, the manner of combination of these variables could well be different with response members. In order to obtain this information, in Exp. I similarity and association values of stimulus members were varied in a 2 × 2 orthogonal design, and in Exp. II similarity and association values of response members were varied in a 2 × 2 orthogonal design.

Drive Variables

Anxiety and stress.—Variations in scores on the Taylor Manifest Anxiety scale have been used to specify levels of drive during the acquisition of stimulus-response relationships of both simple and more complex tasks (Spence, 1958). Verbal statements of failure—failure stress—have been used to induce presumed drive levels. In Exp. I high and low levels of drive defined by high and low scores on the Taylor Manifest Anxiety scale were combined with stress and nonstress. In Exp. II, the levels of drive were defined only by high and low *MA* scores. While the manner of combination of these two specifications of drive was of some interest, the more important concern was the modes of combination of these drive variables with task variables of paired-associate learning.

Task variables and drive variables.—The two most important contemporary formulations of the possible modes of combination of task variables of paired-associate learning and of drive variables are what are here labeled the drive-dominant initial response analysis and the interference analysis. The drive-dominant initial response analysis as developed by Spence predicts that when correct responses to stimulus members of paired-associate units are stronger initially than incorrect responses, more rapid initial increases in performance should occur under high than under low drive. When incorrect responses are stronger initially than correct responses, better initial performance should be obtained under low than under high drive. The interference analysis (Child, 1954; Farber, 1955; Spence, 1958; Taylor, 1958) assumes that increases in drive increase the numbers and strengths of extralist and perhaps intralist responses which compete with correct responses to stimulus members. Accordingly, regardless of the initial relative strengths of correct and incorrect responses, this analysis predicts an inverse relationship between learning speed and drive.

Spence and his associates (Besch, 1959; Spence, 1958; Spence, Farber, & McFann, 1956; Spence, Taylor, & Ketchel, 1956) have constructed lists of paired-associates in which there is initial superiority of correct to incorrect responses or of incorrect to correct responses. While these investigators have obtained better results with the former than with the latter lists, neither Buchwald (1959) nor Kamin and Fedorchak (1957) obtained the predicted reversal.

The more common procedure for the construction of lists of paired-associates is to select stimuli for which there are no or minimal pre-established direct associations between stimulus and response members. Spence has already noted one difference between the two procedures;

when lists have been constructed with no pre-established associations between stimulus and response members, one or more trials must occur before correct or intralist incorrect responses to a stimulus member gains initial dominance. Because of this and other differences, findings with lists constructed by Spence's procedure would be expected to differ from those with lists constructed by the conventional procedure. Thus, what is required is direct determination of the mode of combination of drive variables with characteristics of stimuli of paired-associate lists constructed by the conventional procedure. In Exp. I of this study, the 2×2 orthogonal variations in similarity and association values of stimulus members were combined with the 2×2 orthogonal variations in manifest anxiety and stress. In Exp. II, the 2×2 orthogonal variations of similarity and association values of response members were administered to high-anxiety and low-anxiety *Ss*. (Because too few *Ss* were available at the time this experiment was undertaken, stress was not used.)

With stimulus members and with response members of low similarity and high association values correct responses might soon become dominant over incorrect responses. With stimulus members and with response members of high similarity and low association values, after a few trials, incorrect responses might be temporarily stronger than correct responses. If so, the implication of the drive-dominant initial response analysis is of better performance under high than under low drive with stimulus and with response members of low similarity and high association values. Conversely, with stimulus and with response members of high similarity and low association values poorer initial performance is predicted under high than under low drive. The interference analysis predicts a direct relationship between retardation of performance and drive early and probably late in learning, regardless of similarity and association values of stimulus or of response members.

Experiment I

Method

Paired-associates.—The stimulus members of the four lists of eight pairs of three-letter nonsense syllables were used. The stimulus members were one of four combinations of high or low similarity with high or low association values. Both sets of stimulus members of high similarity were made up of six consonants and three vowels. The Glaze association values of the set with high association values were 93% and 100% ($M = 97.4\%$) and those of the set with low association values were 0% and 7% ($M = 2.6\%$). The syllables of both low-similarity sets were made up of 16 consonants and four vowels. One set had mean Glaze values of 98.2% and the other had mean Glaze values of 2.6%.

The four sets of stimulus members were each paired with the same set of response members whose 15 different letters made them of intermediate similarity. Their 47% and 53% association values ($M = 49.2\%$) were also intermediate.

A modified Saltz and Myers' (1955) group presentation procedure was employed. Each of the eight paired-associates of each of the lists was first reproduced in pica capitals on $4\frac{1}{4} \times 2\frac{3}{4}$ in. sheets of paper. Stimulus members of pairs were on the left and response members were on the right. In order to minimize serial effects, the eight sheets for the paired-associates of each list were assembled in four different random orders. Each of the four orders for each list was then stapled as a booklet in one of the four quadrants of a rectangular (7×11 in.) piece of posterboard. A vertical column of only stimulus members of the paired-associates was typed in pica capitals on each of the $4 \times 5\frac{1}{2}$ in. pages of a booklet on the reverse side. Any one of the four orders in which stimulus and response members appeared on a given trial was always followed by stimulus members alone in the same order.

Procedure for learning.—The four lists were administered simultaneously to groups

of about 10 *Ss*. The part of the instructions which was common for all groups and for all lists explained the design and use of the booklets. Each trial began by telling *Ss* the booklet for that trial. They then waited for *E* to say "begin," whereupon the cover sheet of the appropriate booklet was opened. Each pair of syllables was examined by *Ss* with the intention of later being able to write each of the syllables on the right opposite the syllable on the left with which it had been paired. Four seconds were allowed for the examination of each pair before the signal to turn to the next pair. After viewing the last pair, *Ss* turned their posterboards over and waited for the signal to lift the cover sheet of the booklet on that side and to begin writing the correct responses to stimulus members alone. They had 32 sec. to write those responses after which they tore out that page, wrote their names on it, and passed it to a second *E*. The posterboard was then turned back to the first side. During the next trial the second *E* pretended to be scoring the answer sheet.

Twenty learning trials were administered to all groups with each of the four orders of the pairs of each list occurring five times in an otherwise random sequence.

Procedure for stress.—Stress was defined by two conditions: a statement to *Ss* before beginning that the paired-associate task and academic performance were related, and subsequent statements of failure after Trials 6 and 10. After Trial 6 the second *E* left the table where he was scoring, came in front of the group, and announced, "As you have probably noticed, we have been scoring your test forms. Aren't you people trying? Maybe you're not cooperating. In comparison to similar groups who have taken this test previously, you people are doing significantly poorer. Now, let's try to put in a little more effort." A similar statement was made after Trial 10. The person who served as the second *E* was of mature appearance, considerably above average in height and size, and had a relatively deep, masculine voice.

The same *E* also interrupted nonstress *Ss* after Trial 6 and 10 but with innocuous remarks of about the same lengths as the failure statements. These remarks controlled for the possibility that failure statements merely distracted. Upon completion of the learning task, *Ss* of the stress condition were told the purpose of their experiences and asked not to describe them to anyone else.

Subjects.—The Taylor Manifest Anxiety scale (1953) was administered to students in introductory psychology during the first and second semesters of the 1955–56 academic year. Eighty high-anxiety *Ss* and 80

TABLE 1

MEANS AND *SD*s OF CORRECT RESPONSES FOR TRIALS 1–20 IN EXPERIMENTS I AND II

Treatment Combinations				Experiment I (Stimulus Members)		Experiment II (Response Members)	
Similarity	Association Values	Anxiety	Stress	Mean	*SD*	Mean	*SD*
High	High	High	Yes	92.8	42.5		
High	High	High	No	101.9	26.0	111.8	23.6
High	High	Low	Yes	91.5	39.2		
High	High	Low	No	86.1	26.8	113.9	25.0
High	Low	High	Yes	66.0	36.5		
High	Low	High	No	89.0	25.6	75.9	28.4
High	Low	Low	Yes	61.9	24.7		
High	Low	Low	No	73.8	25.6	76.9	33.5
Low	High	High	Yes	120.6	13.1		
Low	High	High	No	103.0	20.0	125.0	27.6
Low	High	Low	Yes	115.7	20.0		
Low	High	Low	No	124.2	24.5	125.1	18.3
Low	Low	High	Yes	110.3	27.7		
Low	Low	High	No	91.1	25.6	99.5	29.2
Low	Low	Low	Yes	86.1	29.4		
Low	Low	Low	No	99.6	21.9	94.8	36.5

Note.—$N = 10$ in each combination.

TABLE 2

ANALYSIS OF VARIANCE OF MEANS FOR SUCCESSIVE FIVE-TRIAL BLOCKS IN EXPERIMENTS I AND II

Source	Experiment I (Stimulus Members)			Experiment II (Response Members)		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *Ss*	159			79		
Similarity (Sim)	1	5,499.02	25.67**	1	1,357.13	6.11*
Association values (AV)	1	3,900.62	18.21**	1	5,176.15	23.32**
Anxiety (A)	1	200.26	—	1	0.70	—
Stress (St)	1	88.51	—			
Sim × AV	1	4.23	—	1	91.38	—
Sim × A	1	213.91	—	1	18.53	—
Sim × St	1	445.56	2.08			
AV × A	1	182.76	—	1	10.88	—
AV × St	1	187.06	—			
A × St	1	172.22	—			
Sim × AV × A	1	138.75	—	1	4.27	—
Sim × AV × St	1	120.75	—			
Sim × A × St	1	1,113.02	5.20*			
AV × A × St	1	15.62	—			
Sim × AV × A × St	1	1.57	—			
Error (b)	144	214.23		72	221.94	
Within *Ss*	480			240		
Blocks of Trials (T)	3	15,088.05	858.25**	3	6,275.45	—
T × Sim	3	218.34	12.42**	3	6.28	—
T × AV	3	18.84	1.07	3	79.86	4.72**
T × A	3	14.15	—	3	4.93	—
T × St	3	42.95	2.44			
T × Sim × AV	3	20.17	1.14	3	52.24	3.09*
T × Sim × A	3	2.84	—	3	9.94	—
T × Sim × St	3	70.51	4.01**			
T × AV × A	3	8.42	—	3	45.13	2.67*
T × AV × St	3	3.04	—			
T × A × St	3	24.68	1.40			
T × Sim × AV × A	3	64.72	3.68*	3	6.22	—
T × Sim × AV × St	3	5.91	—			
T × Sim × A × St	3	51.04	2.90*			
T × AV × A × St	3	13.57	—			
T × Sim × AV × A × St	3	5.46	—			
Error (w)	432	17.58		216	16.91	

* Significant at .05 level.
** Significant at .01 level.

low-anxiety *Ss* were drawn from among those *Ss* whose manifest anxiety scores were, respectively, in the upper fifth (scores ≥ 20) and lower fifth (scores ≤ 10) of the distribution of all scores. Ten *Ss* from each of these two groups of 80 *Ss* were then assigned randomly to each of the eight combinations of high or low similarity, high or low association values, and stress or nonstress.

Results

Table 1 contains means and *SD*s of numbers of correct responses for all 20 trials for the 16 combinations of stimulus characteristics and drive conditions. Differences among means were assessed by the analysis of variance summarized in Table 2. This analysis includes a breakdown of the means for the 20 trials into means for successive blocks of 5 trials which do not appear in Table 1 but are the Trials (T) variable of the within-*Ss* components in Table 2.

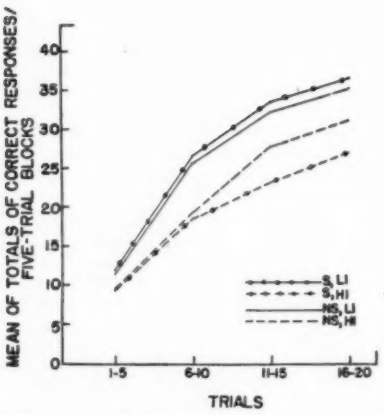

FIG. 1. Means of correct responses in successive blocks of five trials for combinations of stress (S) and nonstress (NS) with lists of high (HI) and low (LI) intralist similarity of stimulus members.

The main effects for similarity (Sim) and association values (AV) were significant beyond the .05 level. Summing over the other variables, the mean of 106.3 correct responses for 20 trials with stimulus members of low similarity was about 30% greater than the mean of 82.9 for stimulus members of high similarity. Stimulus members with high association values (M = 104.5) had nearly as great an advantage over those with low association values (M = 84.7). The Fs for stress, anxiety, and their interactions were not significant.

Stress and similarity entered into three interactions significant at the .01 and .05 levels. The curves in Fig. 1 for the similarity, stress, and trials interaction illustrate the pattern of these interactions. The four curves diverge most during the last 10 trials. During these trials, the curve for low similarity lists rose less rapidly under the nonstress condition than that under stress. Both curves begin and remain above those for the high similarity lists on which stress produced some retardation. In other interactions involving one or more of the other variables with stress and similarity the same pattern held. Stress proved to be facilitative with lists in which stimulus members were of low similarity and inhibitory with lists in which stimulus members were of high similarity.

EXPERIMENT II

Method

Paired associates.—Stimulus and response members of the pairs of the four lists of Exp. I were interchanged to obtain the lists of this experiment. Thus, stimulus members were of intermediate similarity and association values and response members were combinations of high or low similarity and high or low association values. In all other respects, materials for the administration of these lists paralleled those of Exp. I.

Procedure for learning.—Except for the omission of the stress-nonstress condition the procedure for learning the paired-associates of Exp. I was followed.

Subjects.—Eighty new *S*s were drawn from among students in introductory psychology during the second semester of 1956–57. Half had manifest anxiety scores which placed them in the upper fifth of the distribution of those scores; the other half had scores which placed them in the lower fifth. The cutoff scores were those of Exp. I.

Ten of the *S*s with high anxiety and 10 of those with low anxiety were assigned randomly to each of the four combinations of similarity and association values of response members.

Results

Experiment II.—Means and *SD*s of numbers of correct responses for all 20 trials for the eight combinations of characteristics of response members and manifest anxiety appear in Table 1. With the exception of the stress-nonstress variable the analysis of variance for differences among these means was the same as that for Exp. I (Table 2).

The Fs for the main effects of similarity (Sim) and association value

(AV) were, as in Exp. I, significant beyond the .05 level. The mean for correct responses of response members of low similarity ($M = 111.1$) was greater than that for high similarity ($M = 94.6$). Those for high ($M = 119.0$) and low ($M = 86.8$) association values also differed significantly. The main effect for manifest anxiety was not significant, nor did this variable enter into any significant interactions whose patterns were consistent.

Experiments I and II.—The effects of manipulating the stimulus and response sides on the learning of paired-associates were compared by analysis of variance (Table 3). (Manifest anxiety was not included as a factor because it had no apparent effects in Exp. I and, at most, limited effects in Exp. II.)

Significantly more correct responses were made during the 20 trials with variations in response members ($M = 102.9$) than with variations in stimulus members ($M = 94.8$). As expected from the separate analyses, low similarity of stimulus and response members together ($M = 107.9$) led to faster learning than did high similarity of both members together ($M = 86.8$), as was the case for high association values of stimulus and response members together ($M = 109.3$) in relation to low association values together ($M = 85.4$). The difference of 23.4 between means for stimulus members of low ($M = 106.3$) and high ($M = 82.9$) similarity was greater than the difference of 16.5 between means for response members of low ($M = 111.1$) and high ($M = 94.6$) similarity. For stimulus members of high ($M = 104.5$) and low ($M = 84.7$) association values, the difference of 19.8 was less than the difference of 32.2 for response members of high ($M = 119.0$) and low ($M = 86.8$) association values. The suggestion is that similarity had a greater effect on stimulus than on response members; and the association values had a greater effect on response than on stimulus members. Yet, these differences were not sufficient to produce a significant interaction of either similarity or association values with the member varied. Nor was the interaction of both similarity and association values with the member varied significant.

TABLE 3

ANALYSIS OF VARIANCE OF MEANS FOR SUCCESSIVE FIVE-TRIAL BLOCKS FOR SIMILARITY AND ASSOCIATION VALUES OF STIMULUS MEMBERS AND RESPONSE MEMBERS

Source	*df*	*MS*	*F*
Between *Ss*	239		
Response or stimulus member (M)	1	910.25	4.24*
Similarity (Sim)	1	6,693.98	31.22**
Association Value (AV)	1	8,562.17	39.93**
M × Sim	1	162.17	—
M × AV	1	514.60	2.40
Sim × AV	1	51.81	—
M × Sim × AV	1	43.81	—
Error (b)	232	214.41	
Within *Ss*	720		
Blocks of trials (T)	3	21,310.25	1,199.23**
T × M	3	53.25	3.00*
T × Sim	3	164.82	9.28**
T × AV	3	75.36	4.24**
T × M × Sim	3	59.81	3.36*
T × M × AV	3	23.34	1.31
T × Sim × AV	3	57.59	3.24*
T × M × Sim × AV	3	14.82	—
Error (w)	696	17.77	

* Significant at .05 level.
** Significant at .01 level.

DISCUSSION

Similarity and association values.—The slower learning with stimulus members of high similarity than with those of low similarity was consistent with previous findings with nonsense syllables (Underwood, 1953a). However, that this relationship held for similarity of response members and there was no significant difference in the effects of similarity of stimulus and response members were not in accord with Underwood's findings. Underwood used lists

in which similarity of both stimulus and response members were varied; in the present study similarity of either stimulus members or of response members but not of both members was varied. Also, all of his nonsense syllables were of intermediate association values.

The combinations of association values of stimulus and response members of the lists of this study are most comparable to those lists in Mandler and Campbell's (1957) experiment, and in Cieutat, Stockwell, and Noble's (1958) Exp. II. These lists had stimulus members of different association values and response members of intermediate association values or stimulus members of intermediate association values and response members of different association values. For these lists, Mandler and Campbell found that learning speed was related to association values of both stimulus and response members by an inverted U function in their Exp. I and by a U function in their Exp. II. In contrast, the results of this study and those of Cieutat, Stockwell, and Noble suggest a direct monotonic relationship between learning speed and association values of both stimulus and response members. On the whole, the results of other studies suggest that the relationships for association values of stimulus members, response members or both are direct and monotonic (e.g., Feldman & Underwood, 1957; Hunt, 1959; L'Abate, 1959; Noble & McNeeley, 1957; Sheffield, 1946).

The results of this study suggested that association values of stimulus and of response members had essentially the same effects on learning speed. In contrast, Cieutat, Stockwell, and Noble (1958) found that association values of response members had significantly greater effects than association values of stimulus members. Their results are more typical (e.g., Hunt, 1959; Noble & McNeely, 1957; Sheffield, 1946). It is probable that one or more of several differences in the conditions of the present study and the conditions of these other experiments accounts for this difference in results. In the latter experiments, responses were pronounced rather than spelled, or both members of the pairs were of high or low association values rather than one member of intermediate association values, or stimulus or response members of varying association values appeared within the same lists rather than in different lists.

Task variables and drive.—The results of Exp. I suggested with stimulus members of low similarity (easier) stress was facilitative and with stimulus members of high similarity (harder) stress was inhibitory. Thus, whether stress is facilitative or inhibitory depends on the characteristics of the lists. This is a finding which is in accord with implications of the drive-dominant initial response analysis, but not with that of the interference analysis of the role of drive in paired-associate learning.

That stress interacted with similarity of stimulus members but not with their association values supports Sheffield's (1946) proposal of different bases for the effects of similarity and association values on acquisition. High similarity in terms of common physical elements may increase interstimulus generalization sufficiently to bring about initial dominance of intralist incorrect responses rather than of correct responses.

The relatively slower learning which characterizes lists with stimulus members of low rather than of high association values, regardless of their similarity, is probably not based on interstimulus generalization. Responses of repeating stimuli of low association values presumably have longer and more variable latencies and durations and more variable forms than such responses to stimuli with high association values. Such greater variability for both stimulus and response members in stimuli produced by responses of repeating the low than for those of high association values probably accounts for the direct relationships between learning speeds and association values of both stimulus and response members.

There was no indication that anxiety interacted with similarity either of stimulus or of response members. In contrast, Taylor and Chapman (1955)

and Taylor (1958) reported more rapid learning of paired-associates of low intralist similarity and low association values by *Ss* with high than with low anxiety. There is no simple explanation of the lack of agreement between these findings and those of the present study. Differences in the lists may be one factor. Also, the group procedure of administering the paired-associates used here may have been sufficiently similar to classroom test situations to raise the anxiety levels of all *Ss* to about the same high levels, regardless of their manifest anxiety scores.

In Exp. II, the interaction of anxiety, association values, and trials was just significant at the .05 level. Other interactions involving these variables were not significant. Therefore, the results of this study seem best interpreted as consistent with L'Abate's (1959) finding of no interaction of anxiety and association values. Thus, this aspect of the results is not in accord with implications of either the drive-dominant initial response or interference analysis.

Summary

In Exp. I, four lists of eight paired-associates representing 2 × 2 orthogonal variations in similarity (number of letters in common) and association values (Glaze) of nonsense-syllables stimulus members were learned under presumed drive conditions indicated by high or low Taylor Manifest Anxiety scores combined with failure stress or nonstress. The nonsense-syllable response members of these lists were of intermediate similarity and intermediate association values. In Exp. II, the stimulus and response members of the four lists were reversed so that the response members represented 2 × 2 orthogonal variations in similarity and association values. These lists were learned by *Ss* with high or low anxiety scores.

For both experiments lists were administered for 20 learning trials by a group procedure. The failure stress of Exp. I was introduced after Trials 6 and 10.

Learning speeds varied directly with association values of both stimulus and response members, and inversely with similarity of both stimulus and response members. Stress seemingly facilitated acquisition of lists with stimulus members of low similarity relative to those with high similarity and retarded acquisition of lists with stimulus members of high similarity relative to those with low similarity. Stress had no effects on acquisition of stimulus members of high or low association values; anxiety had no effects on acquisition of response members of varying similarity or association values either alone or in combination. The form of interaction of stress with similarity was consistent with Spence's drive-dominant initial response analysis rather than with an interference analysis of the role of drive in complex tasks. However, this was the only finding which supported either analysis.

References

Besch, N. F. Paired-associates learning as a function of anxiety level and shock. *J. Pers.*, 1959, **27**, 116–124.

Buchwald, A. M. Manifest anxiety level, verbal response-strength, and paired-associates learning. *Amer. J. Psychol.*, 1959, **72**, 89–93.

Child, J. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

Cieutat, V. J., Stockwell, F. E., & Noble, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m*, *m'*) in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

Farber, I. E. Motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–325.

Feldman, S. M., & Underwood, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **58**, 11–15.

Gibson, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.

Goss, A. E., Morgan, C. H., & Golin, S. J. Paired-associates learning as a function of percentage of occurrence of response members (reinforcement). *J. exp. Psychol.*, 1959, **52**, 96–104.

Hunt, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, **57**, 262–267.

Jantz, E. M., & Underwood, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, **56**, 174–179.

Kamin, L. J., & Fedorchak, O. The Taylor scale, hunger, and verbal learning. *Canad. J. Psychol.*, 1957, **11**, 212–218.

L'Abate, L. Manifest anxiety and the learning of syllables with different associa-

tive values. *Amer. J. Psychol.*, 1959, **72,** 107–110.

LEVITT, H. Paired-associates learning as a function of manifest anxiety, stress, intra-list similarity, and stimulus association value. Unpublished master's thesis, University of Massachusetts, 1956.

LIFTON, H. An experimental test of the occurrence of intermodal transfer in human verbal learning. Unpublished master's thesis, University of Massachusetts, 1956.

MANDLER, G. H., & CAMPBELL, E. H. Effects of variations in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54,** 269–273.

MANDLER, G. H., & HUTTENLOCHER, J. The relationship between associative frequency, associative ability, and paired-associate learning. *Amer. J. Psychol.*, 1956, **69,** 424–428.

NOBLE, C. E., & MCNEELY, D. A. The role of meaningfulness (*m*) in paired-associate learning. *J. exp. Psychol.*, 1957, **53,** 16–23.

SALTZ, E., & MYERS, T. I. A method for group presentation of paired-associates learning materials. Paper read at Midwestern Psychological Association, Chicago, 1955.

SHEFFIELD, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctoral dissertation, Yale University 1946.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13,** 131–141.

SPENCE, K. W., FARBER, J. E., & MCFANN, H. A. The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning. *J. exp. Psychol.*, 1956, **52,** 296–305.

SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52,** 306–310.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.

TAYLOR, J. A. The effects of anxiety level and psychological stress on verbal learning. *J. abnorm. soc. Psychol.*, 1958, **57,** 55–60.

TAYLOR, J. A., & CHAPMAN, J. P. Anxiety and the learning of paired-associates. *Amer. J. Psychol.*, 1955, **68,** 671.

UNDERWOOD, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intra-list similarity. *J. exp. Psychol.*, 1953, **45,** 133–142. (a)

UNDERWOOD, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intra-list similarity. *J. exp. Psychol.*, 1953, **45,** 143–149. (b)

(Received July 5, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 253–262

THE OBSERVING RESPONSE IN DISCRIMINATION LEARNING[1]

RICHARD C. ATKINSON

Stanford University

Recently, the importance of an orienting or observing response has been emphasized in the formulation of a general theory of discrimination behavior (Atkinson, 1958, 1960; Burke & Estes, 1957; Restle, 1959; Wyckoff, 1952). Unfortunately, for many experimental problems it is not clear how such a theory should be formalized. In particular, there are not enough experimental data available to permit a detailed specification of the postulates relating observing responses and such variables as stimulus dimensions, reinforcement schedules, and stimulus schedules. The purpose of this study is to gain information about this class of relations by modifying the typical discrimination task so that observing responses can be categorized and directly measured.

The experimental situation is considered as a sequence of discrete trials. Each trial is described in terms of the following classifications:

T_1, T_2: *Trial type.* Each trial is either a T_1 or a T_2. Trial type is selected by E and determines in part the stimulus event occurring on that trial.

O_1, O_2: *Observing responses.* At the start of each trial, S makes either an O_1 or O_2. The particular observing response made determines in part the stimulus event for that trial.

s_1, s_e, s_2: *Stimulus events.* Following the observing response, one and only one of these stimulus events (discriminative cues) occurs. On a T_1 trial, s_1 or s_e can occur; on a T_2 trial, s_2 or s_e can occur.

A_1, A_2: *Discrimination responses.* On each trial S makes either an A_1 or A_2 response to the presentation of the stimulus event.

E_1, E_2: *Reinforcing events.* The trial is terminated with the occurrence of one of these events. An E_1 indicates that A_1 was the correct response for that trial and E_2 indicates that A_2 was correct.

The sequence of events on a trial is as follows: (*a*) ready signal occurs and S makes either an O_1 or O_2; (*b*) following the observing response s_1, s_2, or s_e is presented; (*c*) to the onset of s_i, S makes either A_1 or A_2; (*d*) the trial is terminated with reinforcing event E_1 or E_2.

The trial type and reinforcing event are determined by E. The probability of an E_1 event on a T_1 trial is denoted by π_1, and the probability of an E_1 event on a T_2 trial is denoted by π_2. Consequently, the probability of an E_2 is $1 - \pi_1$ on a T_1 trial and $1 - \pi_2$ on a T_2 trial. The two types of trials are equiprobable in the present experiment.

The particular s_i event that is presented on any trial depends on the trial type and the observing response. If an O_1 is made, then (*a*) with probability α the s_1 event occurs on a T_1 trial and the s_2 event occurs on a T_2 trial, and (*b*) with probability $1 - \alpha$ the s_e event occurs, regardless of the trial type. If an O_2 is made, then (*a*) with probability α the s_e event occurs, regardless of the trial type, and (*b*) with probability $1 - \alpha$ the s_1 event occurs on a T_1 trial and the s_2 event occurs on a T_2 trial.

To clarify the experimental procedure, consider a case where $\alpha = 1$, $\pi_1 = 1$, and $\pi_2 = 0$. If S is to be correct on every trial, he must make an A_1 on a T_1 trial and an A_2 on a T_2 trial. However, S can gain information about the trial type only by making the appropriate observing response. That is, O_1 must be made in order to identify the trial type; the occurrence of O_2 always leads to the presentation of s_e. Hence, for perfect responding in this case, S must make the O_1 response with probability 1 and then make A_1 to s_1 or A_2 to s_2.

[1] This research was supported by the Office of Naval Research under Contract Nonr 233(58).

The aim of this study is to investigate the effect of various event schedules on observing behavior. In particular, we are interested in the values of π_1, π_2, and α as determiners of the probability of an O_1 response.

Theory

The analysis of the data will be organized within the framework of a Markov chain model which is closely related to stimulus sampling theory as first formulated by Estes (1950) and Estes and Burke (1953). The mathematical techniques for the model considered in this paper have been presented in detail elsewhere (Atkinson, 1960; Suppes & Atkinson, 1960) and the reader is referred there for a rigorous development.

The basic assumption for observing responses is that if $O_i (i = 1, 2)$ occurs and leads to the selection of a stimulus which in turn elicits a correct discrimination response, then S will tend to repeat that observing response on the next trial. However, if O_i occurs and leads to the selection of a stimulus which elicits an incorrect discrimination response, then S will tend not to repeat that observing response on the next trial. Conceptually, this assumption is similar to that proposed by Wyckoff (1952) and Atkinson (1958).

It is next assumed that S can be described by an ordered four-tuple at the start of trial n where (a) the first member is 1 or 2 and indicates whether O_1 or O_2 will be made on trial n, (b) the second member is 1 or 2 and indicates whether s_1 is conditioned to A_1 or to A_2 (i.e., whether A_1 or A_2 will occur if s_1 is presented), (c) the third member is 1 or 2 and indicates whether s_c is conditioned to A_1 or to A_2, and (d) the fourth member is 1 or 2 and indicates whether s_2 is conditioned to A_1 or to A_2.

These four-tuples will be referred to as subject states and assigned identifying numbers as follows:

1. (1111)	5. (1211)	9. (2111)	13. (2211)
2. (1112)	6. (1212)	10. (2112)	14. (2212)
3. (1121)	7. (1221)	11. (2121)	15. (2221)
4. (1122)	8. (1222)	12. (2122)	16. (2222)

From trial to trial S may change states depending on the sequence of responses and reinforcements. The possible changes are specified by the following axioms:

Axiom 1: With probability θ' the $s_k (k = 1, 2, c)$ stimulus presented on trial n will become conditioned to the reinforced response; if it is already conditioned to that response it remains so. (For example, if s_k is presented and followed by E_j then with probability θ' it will become conditioned to A_j.)

Axiom 2: If $O_i (i = 1, 2)$ is made on trial n and followed by an s_k which elicits a correct discrimination response, then S will repeat the same observing response on the next trial. However, if O_i is made and followed by an incorrect discrimination response, then with probability θ'' S will make the other observing response on the next trial.

From these assumptions and the event schedules employed in this experiment, it can be shown that the sequence of random variables which take the subject states as values is an irreducible, aperiodic Markov chain. This means among other things that a transition matrix $[p_{ij}]$ may be derived from these assumptions where p_{ij} is the conditional probability of being in state j on trial $n + 1$ given state i on trial n. The learning process is completely characterized by these transition probabilities and the initial probability distribution on states.

To clarify the application of the axioms we derive one element of $[p_{ij}]$. Assume S is in State 1211 at the start of Trial n and T_1 E_1 is selected by E with probability $\frac{1}{2}\pi_1$. Then an O_1 occurs with probability 1 and an s_1 is presented with probability α; to the presentation of s_1 an A_2 is made. The S's discrimination response was incorrect and therefore with probability θ'' the observing response changes from O_1 to O_2. Also, with independent probability θ' the conditioning of s_1 changes from A_2 to A_1. Multiplication of the conditional probabilities yields the probability of going from State 1211 to State 2111; i.e., $p_{5,9} = \frac{1}{2}\pi_1\alpha\theta'\theta''$.

In this paper, we shall be primarily interested in the asymptotic behavior of

S. Consequently, $p_{ij}^{(n)}$ is defined as the probability of being in state j on trial $n+1$, given that on Trial 1 S was in State i. Then the following limit exists and is independent of i,

$$u_j = \lim_{n \to \infty} p_{ij}^{(n)}.$$

The quantity u_j can be interpreted as the asymptotic probability of being in state j no matter what the initial distribution. Experimentally, we will be interested in evaluating the following theoretical predictions:

$$P_\infty(O_1) = u_1+u_2+u_3+u_4 +u_5+u_6+u_7+u_8 \quad [1]$$

$$P_\infty(A_1 \mid T_1) = u_1+u_2+u_9+u_{10} +\alpha[u_3+u_4+u_{13}+u_{14}] +(1-\alpha)[u_5+u_6+u_{11}+u_{12}] \quad [2]$$

$$P_\infty(A_1 \mid T_2) = u_1+u_5+u_9+u_{13} +\alpha[u_3+u_7+u_{10}+u_{14}] +(1-\alpha)[u_2+u_6+u_{11}+u_{15}] \quad [3]$$

$$P_\infty(O_1 \bigcap A_1) = u_1+\alpha u_3+(1-\alpha)u_6 +(\alpha/2)[u_4+u_7] +(1-\alpha/2)[u_2+u_5] \quad [4]$$

$$P_\infty(O_2 \bigcap A_1) = u_9+\alpha u_{14}+(1-\alpha)u_{11} +[(1-\alpha)/2][u_{12}+u_{15}] +[1-(1-\alpha)/2][u_{10}+u_{13}] \quad [5]$$

Equation 1 gives the asymptotic probability of an O_1 response. Equations 2 and 3 present the asymptotic probability of an A_1 response on T_1 and T_2 trials, respectively. Finally, Equations 4 and 5 present the asymptotic probability of the joint occurrence of each observing response with an A_1 response.

METHOD

Experimental parameter values.—Six groups of *Ss* were tested. For all groups $\pi_1 = .9$. The groups differed with respect to the experimental parameters π_2 and α; three values of π_2 (.9, .5, and .1) and two values of α (1.00 and .75) were used. Specifically, $\pi_2 = .90$, $\alpha = 1.0$ (Group I); $\pi_2 = .50$, $\alpha = 1.0$ (Group II); $\pi_2 = .10$, $\alpha = 1.0$ (Group III); $\pi_2 = .90$, $\alpha = .75$ (Group IV); $\pi_2 = .50$, $\alpha = .75$ (Group V); and $\pi_2 = .10$, $\alpha = .75$ (Group VI). These particular values of π were selected because they had been used in a similar discrimination experiment where the observing response was not available (Atkinson, Bogartz, & Turner, 1959).

Subjects.—The *Ss* were 240 undergraduates obtained from introductory courses in psychology. They were randomly assigned to groups with the restriction of 40 *Ss* in each group.

Apparatus.—The *Ss* were run in subgroups of two with each S seated in a private booth. The apparatus, viewed from within S's booth, consisted of a shelf at table level which was 30 in. wide and 13 in. deep. A panel 30 in. wide and 30 in. high was mounted vertically on the edge of the shelf farthest from S. Four red panel lights (the s_i stimuli) were in a column and centered on the vertical panel; the bottom light was 20 in. from the base of the panel; the others were spaced above each other at 1¼-in. intervals. Two silent operating keys (the A_1 and A_2 responses) were each mounted 1½ in. in from the edge of the shelf facing S; these keys were 14 in. apart and centered on the column of red lights. On the shelf, 1 in. behind each of these keys, was a white panel light (E_1 and E_2 events). Two additional silent operating keys (the O_1 and O_2 responses) were each mounted 6 in. in from the rear edge of the shelf; these keys were 2 in. apart and also centered on the red lights. A green light (the signal) was centered 3 in. behind the observing response keys on the shelf. The presentation and duration of the lights were automatically controlled.

Procedure.—Within each of the six experimental groups, four subgroups of 10 *Ss* were formed by counterbalancing right and left positions of the observing response and the discrimination response keys. For each S one of the four red lights was randomly designated s_1, another s_e, and another s_2; the fourth light was not used.

The *Ss* were read the following instructions:

> The present study is designed to determine how well you can do on a very difficult pattern recognition problem. We run subjects in pairs to save time, but you are both working on completely different problems. The experiment for each of you consists of a series of trials. The green light on your panel will go on to indicate the start of each trial. Some time later, one or the other of the two lower white lights will go on. Your job is to predict on each trial which one of the two white lights will go on and to indicate your prediction by pressing one of the two lower keys.

However, before you make your prediction you will receive additional information. That is, as soon as the green light goes on, press one or the other of the two upper keys—which key you press is up to you. Shortly thereafter, one of the four red lights will go on. The particular red light which goes on depends in part on the key you press. Further, the red light which goes on will help you in making your prediction as to which white light goes on. After you have seen one of the red lights go on, you will then predict which white light will go on by pressing the proper key. That is, if you expect the left white light to go on, press the left lower key, and if you expect the right white light to go on, press the right lower key. If the light above the key you pressed goes on, your prediction was correct, but if the light above the key opposite from the one you pressed goes on, you were incorrect and should have pressed the other key. Thus, for a single trial, the sequence of events is as follows: (1) the green light goes on to signal the start of the trial, (2) you press one of the two upper keys, (3) one of the red lights will go on, (4) you press one of the two lower keys, (5) if the white light goes on above the key you pressed, your prediction was correct; if the light above the key opposite from the one you pressed goes on, you were incorrect and should have pressed the other key.

Questions were answered by paraphrasing the appropriate part of the instructions. Following the instructions, 200 trials were run in continuous sequence. This sequence was followed by a 5-min. rest period; during this period no questions referring to the experiment were answered by E, and Ss were not allowed to discuss the experiment. Following the rest, 200 additional trials were run. For each S, random sequences of s_i and E_i events were generated in accordance with assigned values of π_1, π_2, α, and the observed O_i responses.

On all trials, the signal light was lighted for 2 sec. The appropriate s_i stimulus light immediately followed the cessation of the signal light and remained on for 3 sec. After

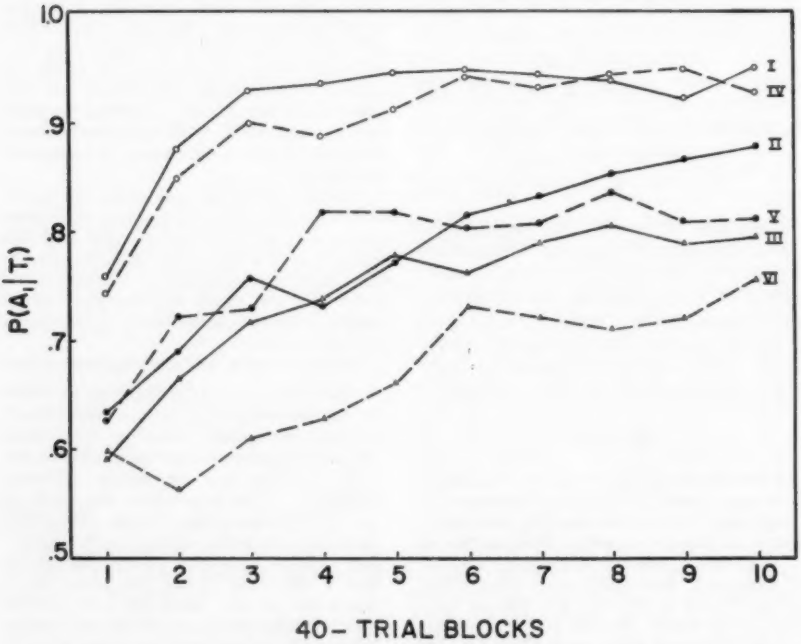

FIG. 1. The average proportion of A_1 responses on T_1-type trials in successive blocks of 40 trials.

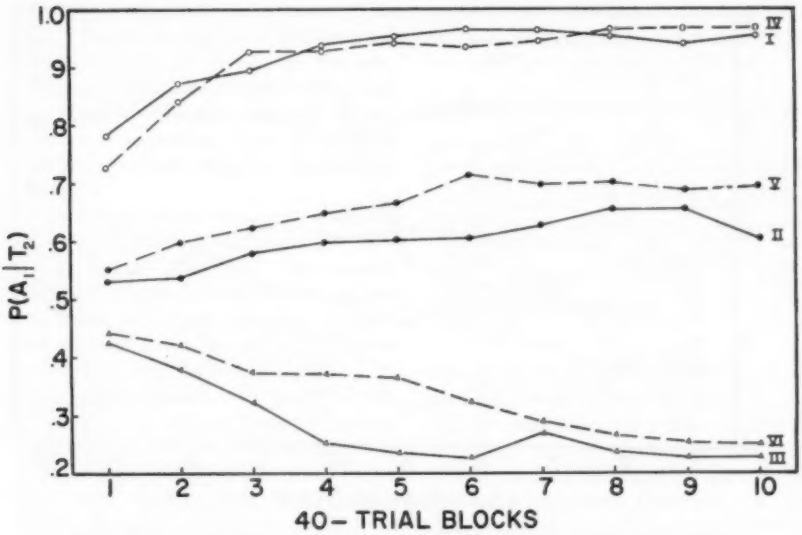

FIG. 2. The average proportion of A_1 responses on T_2-type trials in successive blocks of 40 trials.

the offset of the s_i light, one of the reinforcing lights went on for 2 sec. The time between the offset of the reinforcing light and the onset of the signal light for the next trial was 3 sec.

RESULTS AND DISCUSSION

Mean learning curves and asymptotic results.—Figure 1 presents the average proportion of A_1 responses on T_1-type trials in successive blocks of 40 trials. For each S the proportion of A_1's on T_1 trials was tabulated for a 40-trial block, and these quantities were then averaged over Ss. Similarly, Fig. 2 presents the average proportion of A_1's on T_2-type trials in successive blocks of 40 trials. Finally, Fig. 3 presents the average proportion of O_1 responses. In all three figures the curves appear to be reasonably stable over the last half of the experiment. Consequently, the proportions computed over the final block of 160 trials were used as estimates of asymptotic performance.

Table 1 presents the observed mean proportions over the last 160-trial block and the related SDs. The observed values of $P_\infty(A_1|T_i)$ were computed as indicated in the description of Fig. 1 and 2. The observed values of $P_\infty(O_i \bigcap A_1)$ were computed by obtaining, for individual Ss, the proportion of trials on which both the A_1 response and the O_i response occurred and then averaging over Ss.

The values predicted by the model are also presented in Table 1 for the case where $\theta' = \theta'' = \theta$. Expressions for the u_k's were derived by standard methods (Feller, 1957), and then combined by Equations 1–5 to predict the response probabilities. The computations were performed at the Western Data Processing Center on

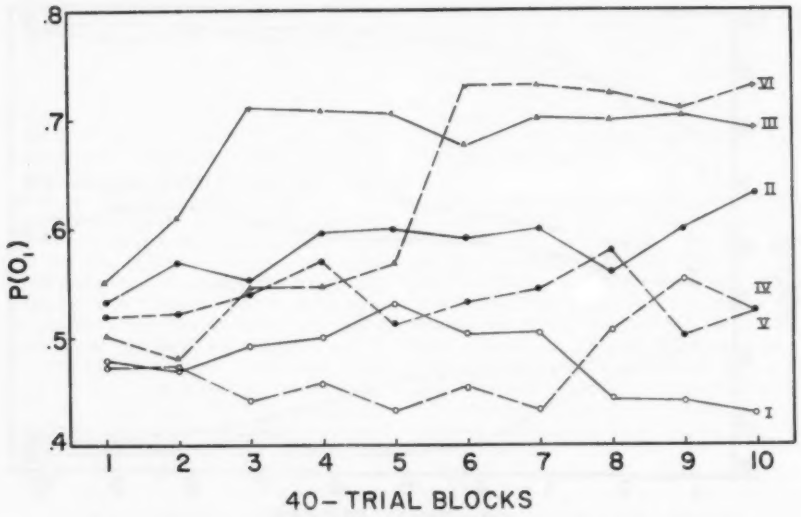

FIG. 3. The average proportion of O_1 responses in successive blocks of 40 trials.

TABLE 1

PREDICTED AND OBSERVED RESPONSE PROBABILITIES OVER THE LAST BLOCK OF 160 TRIALS

	Group I			Group II			Group III		
	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*
$P_\infty(A_1 \mid T_1)$	.90	.94	.014	.81	.85	.164	.79	.79	.158
$P_\infty(A_1 \mid T_2)$	.90	.94	.014	.59	.61	.134	.21	.23	.182
$P_\infty(O_1)$	.50	.45	.279	.55	.59	.279	.73	.70	.285
$P_\infty(O_1 \cap A_1)$	.45	.43	.266	.39	.42	.226	.37	.36	.164
$P_\infty(O_2 \cap A_1)$	.45	.47	.293	.31	.31	.232	.13	.16	.161
	Group IV			Group V			Group VI		
	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*
$P_\infty(A_1 \mid T_1)$	.90	.93	.063	.80	.82	.114	.73	.73	.138
$P_\infty(A_1 \mid T_2)$	.90	.95	.014	.60	.68	.114	.27	.25	.138
$P_\infty(O_1)$	.49	.50	.257	.52	.53	.305	.63	.72	.263
$P_\infty(O_1 \cap A_1)$	.44	.47	.241	.35	.38	.219	.32	.36	.138
$P_\infty(O_2 \cap A_1)$	.46	.47	.247	.34	.36	.272	.19	.13	.168

an IBM 709 computer.[2] By presenting a single value for each theoretical quantity in Table 1 we imply that these predicted proportions are independent of θ. Actually this is not always the case. However for the event schedules employed in this experiment the dependency of the

[2] The program or punch program deck is available to anyone interested in generating theoretical results for parameter values not considered in this paper.

theoretical proportions on θ is negligible. For θ ranging over the interval .00001 to 1.0 the values of the predicted proportions are affected in only the third or fourth decimal place; it is for this reason that we present theoretical values to only two decimal places.[3]

In view of these comments it should be clear that the predictions in Table 1 are based solely on the experimentally assigned values of π_1, π_2, and α. Thus, they are entirely a priori and do not make use of any parameters evaluated from the data. Consequently, differences between *S*s, which can be represented by inter-*S* variability in θ, do not substantially affect these asymptotic predictions. Of course, this implies that the observed proportions for individual *S*s and also proportions averaged over *S*s should both approach these predicted values with increasing sample size.

An inspection of Table 1 indicates good agreement between observed and predicted quantities. The observed value of $P_\infty(A_1|T_1)$ decreases from Groups I to III and also from Groups IV to VI as predicted by the model. Similarly the observed values of $P_\infty(A_1|T_2)$ decrease from I to III and from IV to VI as expected. Column comparisons are also in the appropriate order, that is, on this measure Group I is less than IV, II is less than V, and III is less than VI. Thus, as we increase the frequency of reinforcing the A_2 response on T_2 trials, we not only observe an increment in $P_\infty(A_2|T_2)$ but also a decrement in $P_\infty(A_1|T_1)$.

For $P_\infty(O_1)$, an increase occurs from Groups I to III and from Groups IV to VI in accordance with theoretical results. That is, the proportion of O_1 responses increases as a function of the difference between π_1 and π_2; of course, this result would be expected in view of the fact that differential reinforcement for the observing responses depends on the difference between the reinforcement schedules on T_1 and T_2 trials. However, column comparisons on the $P_\infty(O_1)$ measure for I–IV and III–VI are in the reverse order; the difference on the $P_\infty(O_1)$ measure is particularly large for Group VI. This difference between data and theory for Group VI is also reflected in $P_\infty(O_2 \bigcap A_1)$; in fact, the discrepancies of these two quantities from predicted values are greater than any of the others in Table 1.

TABLE 2

ANALYSIS OF VARIANCE OF THE NUMBER OF O_1 RESPONSES IN THE LAST BLOCK OF 160 TRIALS

Source	*df*	*MS*	*F*
π (values of π_2)	2	27,533.2	13.8*
α values	1	113.4	0.1
O (O_1 left or right)	1	22,253.0	11.1*
A (A_1 left or right)	1	44.2	0.0
Interactions (11)	18		—[a]
Within	216	2,000.7	

[a] None significant at .05 level.
* $P < .001$.

An analysis of variance on the number of O_1 responses in the last block of 160 trials is presented in Table 2. The effects of the O_1 and A_1 placements on *S*'s panel (i.e., right or left) are included in the analysis. The effect of the π-variable is highly significant as would be expected. However, the α-variable is not significant. This finding might have been anticipated since the theoretical prediction for the over-all effect of α is small for the parameter values used in this study. The most unexpected result of the analysis is with regard to the observing response

[3] Essentially the same statement holds for $\theta' \neq \theta''$. However, in some cases the dependency is slightly larger.

variable; the placement of the O_1 key turns out to be highly significant while the placement of the A_1 key has no effect. Over all groups and *S*s for the last 160 trials, the right hand observing response key was chosen on 55% of the trials while the right hand A_j key was selected on 50% of the trials. This right position preference on the observing response keys is particularly surprising in view of the fact that no similar preference exists for the A_j key. Several variables may account for this finding; for example, the observing response keys are in juxtaposition while the A_j keys are well separated; also, the observing response keys are further from *S* than the A_j keys.

In order to evaluate statistically the adequacy of the present model we have run a test suggested by Pillai and Ramachandran (1954) on the $P_\infty(O_1)$ measure. The test involves taking the largest absolute difference between an observed mean value and the predicted value in a collection of samples (in this case six). This difference is then divided by an over-all estimate of the standard error of the mean, that is, it is assumed that the observations are randomly selected from populations with homogeneous variance. As noted above, the largest discrepancy on the $P_\infty(O_1)$ measure occurs for Group VI. The predicted number of O_1s in the last block of 160 trials was 100.8 and the observed mean value was 115.7. The within-cells term in Table 2 was used to estimate the standard error of the mean, and in terms of Cochran's test there was no reason to reject the assumption of homogeneous variance. The obtained value of the Pillai-Ramachandran statistic was 2.1 and was not significant at the .05 level. Consequently, in terms of this particular statistical criterion there is no evidence to suggest that we reject the present model.

As noted earlier, not only group means but also the responses of individual *S*s should approach the theoretical values presented in Table 1. A check on the correspondence between individual asymptotic behavior and predicted values is equivalent to evaluating the agreement between observed *SD*s presented in Table 1 and asymptotic variability predicted by the model. Unfortunately direct computation of the theoretical *SD* is extremely cumbrous, and we have not obtained an analytical result. However, research reported by Suppes and Atkinson (1960) dealing with a similar model found that observed *SD*s were substantially larger than predicted values. Considering the rather large *SD*s reported here, their finding may be applicable to this set of data.

Transition characteristics.—A basic assumption in the model requires that if *S* is correct on trial n (i.e., A_1-E_1 or A_2-E_2 occurs) then on trial $n + 1$ he will repeat the observing response made on trial n. However, if *S* is incorrect (i.e., A_1-E_2 or A_2-E_1 occurs) then with probability θ'' he will shift observing responses from trial n to $n + 1$. This is a strong assumption and yields a highly deterministic set of predictions; for example, repetition of an observing response with probability 1 if *S* is correct on the preceding trial. On the other hand, a weaker form of the assumption which requires only a greater probability of observing response alternation following trials on which incorrect as compared to correct discrimination responses occur seems to be a reasonable conjecture for this type of problem. To test this class of assumptions we have computed the proportions of observing response alternations conditionalized on correct and incorrect discrimination responses over the last 160 trials. Let $N_n(s|c)$ denote the number of *S*s who were correct on trial $n - 1$ and shifted observing responses from trial $n - 1$ to n; also, let $N_n(c)$ be the number of *S*s who were correct

on trial n. Similarly, define $N_n(s|\bar{c})$ and $N_n(\bar{c})$ in terms of incorrect responses. Further, define

$$N(s|c) = \sum_{n=241}^{400} N_n(s|c)$$

and

$$N(c) = \sum_{n=240}^{399} N_n(c)$$

and define $N(s|\bar{c})$ and $N(\bar{c})$ similarly. Then estimates of the conditional probabilities of shifting observing responses following correct or incorrect discrimination responses are, respectively,

$$\hat{P}(s|c) = \frac{N(s|c)}{N(c)},$$

$$\hat{P}(s|\bar{c}) = \frac{N(s|\bar{c})}{N(\bar{c})}$$

Table 3 presents the observed data for each of the groups. No statistical test is needed to see that these observed transition probabilities differ significantly from theoretical values. It suffices to note that theoretically $\hat{P}(s|c)$ should be identically zero for all groups whereas the observed values of this quantity differ markedly from zero. Without regard to the specific assumption considered in this paper, the question can be raised as to whether or not shifting of an observing response is more likely following incorrect or correct trials, that is whether $\hat{P}(s|\bar{c})$ is greater than $\hat{P}(s|c)$. A formal test of this hypothesis is a complex matter and we do not attempt it here. However note that for five of the six groups $\hat{P}(s|\bar{c})$ is greater than $\hat{P}(s|c)$. Further, the difference between these quantities increases as π_2 decreases; that is, the difference increases from Groups I to III and from Groups IV to VI.

TABLE 3

TRANSITION FREQUENCIES AND ESTIMATED PROPORTIONS OVER THE LAST BLOCK OF 160 TRIALS

	Groups					
	I	II	III	IV	V	VI
$N(s\|c)$	1610	880	933	1617	883	735
$N(c)$	5388	4080	4664	5463	4032	4350
$N(s\|\bar{c})$	275	520	446	266	517	524
$N(\bar{c})$	973	2280	1696	897	2328	2010
$\hat{P}(s\|c)$	.299	.216	.200	.296	.219	.169
$\hat{P}(s\|\bar{c})$	.283	.228	.263	.297	.222	.261

In conclusion, the rather striking correspondence between theoretical and observed values in Table 1 lends considerable support to the main features of the model. For the type of discrimination problem considered in this paper, it seems clear that asymptotic behavior can be predicted with accuracy in terms of the particular relations we have postulated among reinforcement schedules, observing responses, and discrimination responses. However, the sequential data reported in Table 3 indicate that some of the detailed features of the stimulus sampling process assumed in the model are certainly incorrect; this finding is not too surprising in view of related research on similar Markov chain models. Fortunately, within the framework of stimulus sampling theory, one can restate our axioms in only slightly modified form and thereby avoid the completely deterministic predictions made by the present model for sequential data. The disadvantage of such a reformulation is that the mathematical complexity of the model is greatly increased. The reader interested in details of such modifications is referred to Suppes and Atkinson (1960).

SUMMARY

An analysis of observing responses in discrimination learning was made. The typical discrimination task was modified so that two mutually exclusive and exhaustive observing responses could be identified and directly recorded. The experimental situa-

tion involved a series of 400 trials, each trial belonging to one of two types (T_1 or T_2). The sequence of events on a trial was as follows: (*a*) ready signal to which *S* made an observing response; (*b*) the presentation of one of three stimuli; (*c*) occurrence of one of two discrimination responses to the stimulus presentation; (*d*) termination of the trial with the reinforcement of a discrimination response. The particular stimulus presented on a trial depended on the observing response and the trial type. Following one of the observing responses, different stimuli were presented on T_1 and T_2 trials so that it was possible for *S* to identify the trial type; following the other observing response, the same stimulus was presented on both types of trials and hence *S* could not identify the trial type.

Six groups of college students were tested. The major independent variable specified different pairs of reinforcement schedules for the two trial types. The results indicated a highly predictable relation between the selection of observing responses and reinforcement schedules. In general, the greater the difference between the reinforcement schedules on T_1 and T_2 trials, the greater the preference for one observing response over the other. The analysis of the data was in terms of a Markov chain model which is closely related to stimulus sampling theory. There was excellent agreement between theoretical and observed values on asymptotic measures of observing and discrimination responses. However, an analysis of the sequential data indicated certain difficulties with the model.

REFERENCES

ATKINSON, R. C. A Markov model for discrimination learning. *Psychometrika*, 1958, **23**, 309–322.

ATKINSON, R. C. A theory of stimulus discrimination learning. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *First Stanford symposium on mathematical methods in the social sciences*. Stanford: Stanford Univer. Press, 1960. Ch. 15.

ATKINSON, R. C., BOGARTZ, W. H., & TURNER, R. N. Supplementary Report: Discrimination learning with probabilistic reinforcement schedules. *J. exp. Psychol.*, 1959, **57**, 349–350.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

FELLER, W. *An introduction to probability theory and its applications*. (2nd ed.) New York: Wiley, 1957.

PILLAI, K. C. S., & RAMACHANDRAN, K. V. On the distribution of the ratio of the i^{th} observation in an ordered sample from a normal population to an independent estimate of the standard deviation. *Ann. math. Statist.*, 1954, **25**, 565–572.

RESTLE, F. A survey and classification of learning models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Ch. 20.

SUPPES, P., & ATKINSON, R. C. *Markov learning models for multiperson interactions*. Stanford: Stanford Univer. Press, 1960.

WYCKOFF, L. B., JR. The role of observing responses in discrimination behavior. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received July 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 263–271

INTERACTION OF LACK OF SLEEP WITH KNOWLEDGE OF RESULTS, REPEATED TESTING, AND INDIVIDUAL DIFFERENCES[1]

ROBERT T. WILKINSON

Applied Psychology Research Unit, Cambridge, England

It has been established (Pepler, 1959; Wilkinson, 1958, 1959) that 30 hr. loss of sleep causes a decline in efficiency towards the end of 25 min. continuous work on Leonard's 5-Choice Test of Serial Reaction, a task in which a series of repeated simple choices have to be made as quickly and accurately as possible. These studies also suggested that this effect of lack of sleep changes little when the environmental warmth is increased and when periodic rest pauses are introduced. These negative results tell us a little about the nature and action of the stress but, perhaps, much more might be learned if we could find factors which influence its effect positively, either increasing or decreasing it. The present experiment has continued the search by examining whether the effect of lack of sleep changes, firstly as we increase feedback of knowledge of results (KR) in the test and, secondly, as *S* becomes more familiar with the test and with the stress. A final question was whether *S*s are consistent in the degree to which they are affected by lack of sleep. The two last-named issues were examined by having each *S* repeat an original 2-wk. program of testing twice so that their tests covered a period of 6 wk. Performances after sleep and no sleep on both an individual and a group basis could then be compared over successive, identical 2-wk. periods.

[1] The Royal Navy supported this work by contributing *S*s, administrative help, and financial assistance through the Royal Naval Personnel Research Committee of the Medical Research Council of the United Kingdom. The work owes much to the influence of N. H. Mackworth and D. E. Broadbent, successive Directors of the Applied Psychology Research Unit where it was carried out.

METHOD

Subjects.—The *S*s were 12 enlisted men between the ages of 18 and 30 yr. All were volunteers to undergo sleep deprivation. They were quartered near the laboratory for 6 wk. to act as *S*s in this experiment, and during this period they lived and worked together as a close social group.

Apparatus and procedure.—The apparatus and form of the 5-Choice Test of Serial Reaction has been described fully by Leonard (1959) and Wilkinson (1958). Briefly, there are five metal contacts and five associated bulbs. At the start of the test one of the bulbs is lighted and *S* taps the corresponding disc with a stylus; this causes the bulb to go out and another to light up, whereupon *S* again taps the appropriate disc, and so the cycle continues. The order in which the bulbs light up over successive taps is prearranged to appear random, and the task is merely to go on tapping the contact whose light comes on as quickly and as accurately as possible for 30 min. Scores are registered on three counters whose readings are recorded by *E* at the end of each 5-min. period of the test. One counter gives the number of correct taps made, another counter records an error whenever *S* taps the wrong disc; a third counts the number of gaps or occasions when *S* allows a period of 1.5 sec. or more to elapse between successive taps.

The variable of KR.—The test can take two forms, one with very complete feedback of KR and one where this feedback is at a minimum. In the former, *S* hears a bellnote whenever he makes an error and a buzz as soon as 1.5 sec. have elapsed since his last response, i.e., when he has scored a gap. Every 5 min. he hears *E*'s voice through a loud-speaker tell-

TABLE 1

DISTRIBUTION OF CONDITIONS AMONG TEAMS AND OVER DAYS IN EACH OF THREE 2-WK. PERIODS

Team	Day 1 (Tuesday)	Day 2 (Thursday)	Day 3 (Tuesday)	Day 4 (Thursday)
A (3 *Ss*)	S:KR	NS:KR	NS:NKR	S:NKR
B (3 *Ss*)	S:NKR	NS:NKR	NS:KR	S:KR
C (3 *Ss*)	NS:KR	S:KR	S:NKR	NS:NKR
D (3 *Ss*)	NS:NKR	S:NKR	S:KR	NS:KR

Note.—S:KR =sleep—knowledge of results; S:NKR =sleep—no knowledge of results; NS:KR =no sleep—knowledge of results; NS:NKR =no sleep—no knowledge of results.

ing him his scores of corrects and errors for the previous 5 min. work, and relating these to scores in the penultimate 5 min. by the comments, "Better," "Worse," or "No change." On the day following the test the full scores are published for all *Ss* to see. In the other form of the test nothing is done to give *Ss* any information on their performance. They hear *E*'s voice every 5 min., but he merely recites a short series of meaningless letters; about the same number of bells and buzzes are heard as in the comparable tests under full KR, but here they bear no relation to the incidence of errors and gaps and *S* knows this; finally, no results are published.

On the day preceding the first test *Ss* were fully acquainted with the procedures described above and were also given 5 min. practice on each form of the test.

The variable of lack of sleep.—The method of sleep deprivation has been described fully elsewhere (Wilkinson, 1958). Tests after loss of sleep were always carried out at the same time of the day as the control tests. They differed from the latter only in that they employed *Ss* who had foregone normal sleep the night before; these *Ss*, staying awake in groups of three, were allowed to amuse themselves, under supervision, in their recreation room. No tests were given during the night.

Design.—The tests were carried out in four conditions: with KR after normal sleep, with KR after no sleep the previous night, with no KR after normal sleep, and with no KR after no sleep. The *Ss* were tested on the Tuesday and Thursday of six consecutive weeks and by the end of the first 2 wk. each one had met each of the four conditions once in an order prescribed by his team as shown in Table 1. The second and third 2 wk. of the 6-wk. program were for each *S* an identical repetition of the first.

RESULTS

Figure 1 is a comprehensive representation of the results, averaged over all 12 *Ss*, in terms of gaps, corrects, and errors. In each of these categories three graphs are shown, one for each 2-wk. period; these show the average score in the four conditions over each 5-min. period of the test.

No sleep and no KR.—Figure 1 shows that these main effects were large and that they varied along the time-base of individual tests and along that of the 6-wk. program of tests. While this influence of time is important it may be simpler to ignore it for the present and assess these main effects on the basis of the total scores in the four conditions over the whole 6 wk. of testing. During this period each *S* met each combination of sleep or no sleep with KR or no KR three times. The difference between his scores in the sleep and no sleep tests will give an index of what we will call his "no-sleep impairment" or the degree to which his performance was impaired as a result of losing sleep. Dealing similarly with the tests with and without KR we can obtain an index of a "no KR impairment." These indices are shown in Table 2 derived from total scores of gaps, corrects, and errors in the four condition-combinations. Their significance is assessed by the nonparametric procedure of Wilcoxon (1949). Table 2 also shows the interaction of lack of sleep with no KR. This is obtained by taking the difference between no-sleep impairments in the tests with and without KR, the significance being assessed as before.

Lack of sleep caused a reduction in corrects ($P < .01$) and an increase in gaps ($P < .01$). Errors were increased also but not significantly.

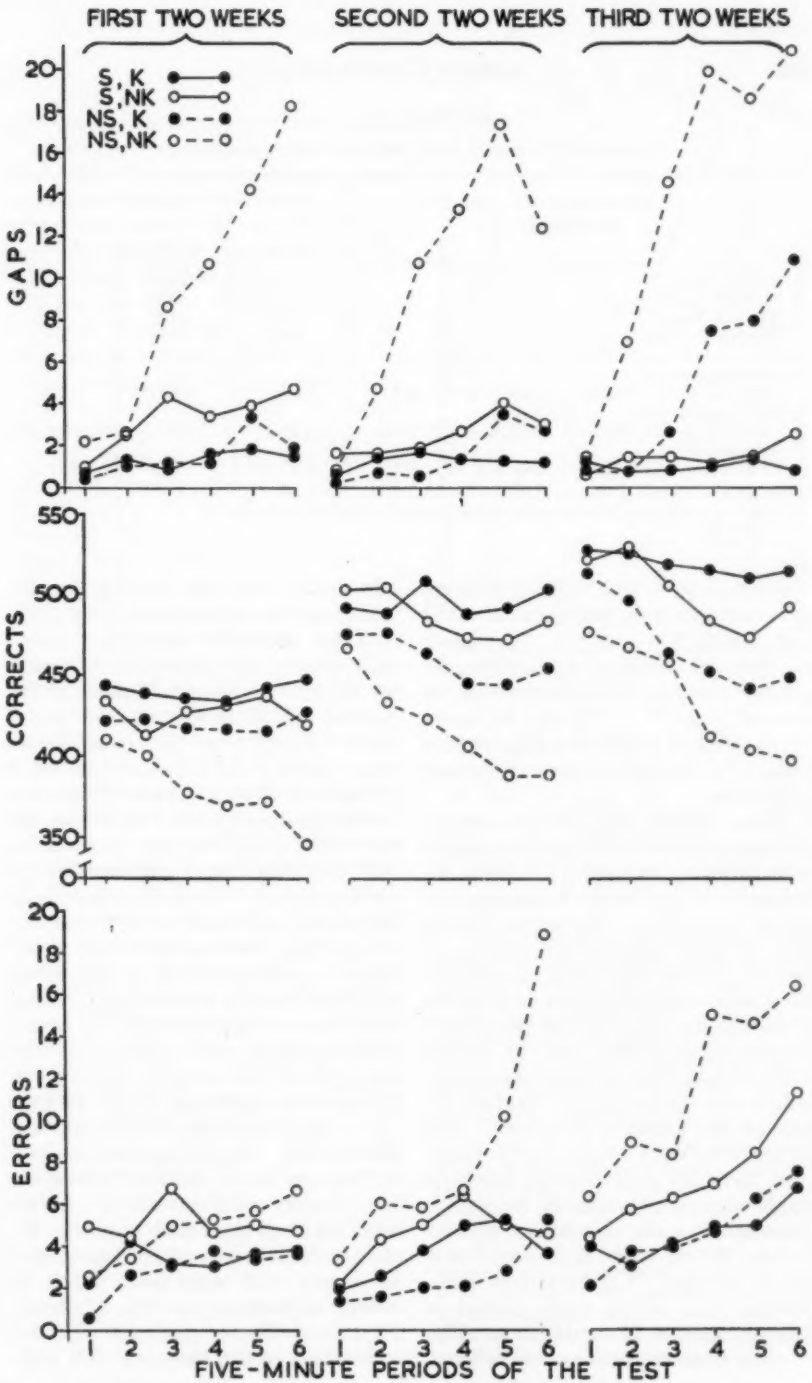

FIG. 1. The effect of combination of sleep and no-sleep with KR and no KR on gaps, corrects, and errors in three successive and identical 2-wk. periods of testing.

TABLE 2

EFFECT OF NO SLEEP (NS), KR, AND THEIR INTERACTION

Conditions	Scores Per Test Averaged Over All *Ss*		
	Gaps	Corrects	Errors
S:KR Tests	7.1	2906	22.6
S:NKR Tests	14.3	2827	33.5
NS:KR Tests	16.1	2696	20.1
NS:NKR Tests	66.3	2464	49.3
NS Impairment: (NS:KR + NS:NKR) − (S:KR + S:NKR)	+61.0**	−573**	+13.3
NK Impairment: (S:NKR + NS:NKR) − (S:KR + NS:KR)	+57.4**	−311*	+40.1**
Interaction: (NS:NKR − S:NKR) − (NS:KR − S:KR)	+43.0**	−153	+18.3*

Note.—Impairments and positive interactions appear as negative in corrects.
* $P < .05$.
** $P < .01$.

The stress of no KR reduced corrects ($P < .02$) and increased gaps ($P < .01$) and errors ($P < .001$). The effect of lack of sleep was significantly greater when no KR was given in the case of gaps ($P < .01$) and of errors ($P < .05$). A similar but insignificant trend ($P < .1$, one-tailed test) appeared in corrects.

Time during the 30-min. test.—Having established the pattern of the main trends, we can now turn to changes in performance during the course of the test. The scores during each 5-min. period are shown in Fig. 1, but for purposes of analysis the test was divided into two halves to obtain an index of the decrement in performance from first to second half. When this is done the scores reflect the same general picture as appears in Table 2 for overall test scores (i.e., first+second half). However there are differences in levels of significance which should be noted. Gaps show higher significance on two counts, the no-sleep impairment and the interaction. Corrects show little change, but errors again reveal a different picture of significance. The no-sleep impairment is now significant ($P < .02$) while the no KR impairment and the interaction cease to be.

There are other interesting points arising from comparison of the trends in the first and second halves of the test of which three only are noted here. Firstly, as has been shown many times before, the adverse effect of lack of sleep on performance becomes greater as the test wears on; the same holds for the stress of no KR, a finding not so well known from previous work. Secondly, there is an interesting contrast to be made between gaps and corrects; the latter show a significant ($P < .01$) effect of lack of sleep in the first half (indeed the effect is significant in the first 5 min.); gaps do not. And yet by the last half of the test the no-sleep impairment is highly significant in gaps ($P < .001$), more so than in corrects ($P < .005$). A puzzling situation prevails in errors; in the first half of the test, performance was slightly better with no sleep than with sleep but by the second half a no-sleep impairment in errors had appeared which is nearly significant at the .05 level. In spite of this, a significant ($P < .02$) interaction of the sleep and KR vari-

ables occurred in the first half which was reduced to insignificance in the last half. Presumably loss of sleep increases errors in the first half only when associated with no KR; by the last half the effect has spread to the tests where KR was given.

Time during the program of tests.—Changes during the course of the 12 tests performed by each *S* over a 6-wk. period will be considered only as they result from lack of sleep. Readers interested in variation of the no KR impairment and the interaction can distinguish these from the graphs in Fig. 1. Similarly, they may examine changes in the main trends across 2-wk. periods as a function of the deterioration in performance during the course of the test.

The no-sleep impairment increased from first to third 2 wk., significantly in the case of errors ($P < .005$) and of gaps ($P < .05$), but insignificantly in corrects, only 8 out of 12 *S*s showing the effect. The Wilcoxon test was again used to assess these levels of significance on a two-tailed basis; this time it was based on the difference in the no-sleep impairments occurring in the first and third 2 wk. These trends are shown in Fig. 2. It will be seen that the main increase has occurred by the second 2 wk. in corrects and errors but not until the third in the case of gaps. Errors were actually fewer after no sleep in the first 2 wk. Similar temporal trends were apparent during the course of the individual tests.

Consistency of Ss in no-sleep impairment.—The no-sleep impairment varied greatly between *S*s. If this index is used as a basis for ranking *S*s in each 2-wk. period then these rankings differ little from one 2 wk. to the next; those men who showed a large effect of lack of sleep in the first did the same in the second and third. This effect was significant in errors ($P < .05$) and corrects ($P < .05$) but not in gaps ($P < .2$)[2]. This consistency suggests that there are indeed real differences in the extent to which individuals are affected by lack of sleep.

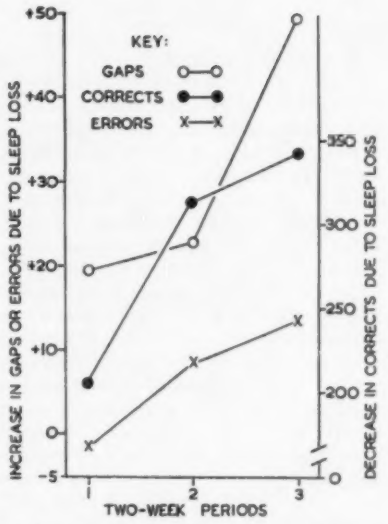

FIG. 2. Change in the impairment of performance due to lack of sleep over three successive and identical 2-wk. periods.

[2] The significance of these results was assessed by calculating Kendall's coefficient of concordance of the rankings in each team of 3 *S*s across 2-wk. periods. The method is described by Siegel (1956, p. 229). Probability values of the coefficients of concordance for each team were then combined to give an overall level of significance of the concordance of the rankings of all 12 *S*s across 2-wk. periods; this was thus independent of any concordance between teams due to their being consistently treated differently, i.e., with respect to the order in which they were given the four conditions (see Table 1). This is a relatively insensitive measure of significance being based only on 8 *df*; that .05 levels of significance were reached in spite of this encourages the view that in an experiment designed solely to investigate this question the same results would appear at much higher levels of significance.

Discussion

It is known that 30 hr. loss of sleep impairs performance in Leonard's 5-Choice Test of Serial Reaction. There is no need to restate the implications of this result; instead we have to discuss firstly how the effect varies as a change is made in the amount of incentive as determined by the feedback of KR. This question was examined in an earlier pilot experiment (Wilkinson, 1957, 1958); when the test was changed by adding KR the effect of lack of sleep upon it was substantially reduced. Some doubt was cast on this result when Williams, Lubin, and Goodnow (1959) found a relatively small and insignificant effect of letting *S* see his successive reaction times registering on a timer during a very simple test of serial reaction (which they rather challengeably describe as a vigilance task). One of the objects of the present experiment therefore was to check our earlier conclusion; it has been confirmed at a high level of confidence. The failure of the result of Williams et al. (1959) to reach significance may be due to the moderate nature of their feedback of KR, a possibility they themselves recognize.

We know that KR tends to raise the level of performance (Elwell & Grindley, 1938) and one reason for this is thought to be that it raises the level of motivation. Now we have seen that this effect is disproportionately large if *S*s have lost sleep and this suggests that in this state the level of motivation is unduly low, giving KR, so to speak, more to work on. Thus we now have some experimental evidence for the suggestion often made that lack of sleep produces a form of fatigue which does not directly cause a work decrement but raises the threshold at which certain work motives are effective (Whiting & English, 1925).

But what is motivation? A series of papers most of which have been mentioned in reviews by Malmo (1957, 1959), have sought to wed it to the concept of arousal. Although it must be conceded that the courtship has had its setbacks there certainly seems to be some relation between the level of activation or arousal of the body and level of performance (Bartoshuk, 1955; Surwillo, 1956), and with motivational variables which may influence this, for example, KR (Stennett, 1957), electric shock (Schnore, 1959), and interest content of a piece of work (Bartoshuk, 1956; Wallerstein, 1954). The present result provides further evidence for this by showing that a motivating influence (KR) can cancel the effects of an influence which according to the EEG findings of Bjerner (1949) and Armington and Mitnick (1959) makes for lowered arousal, namely lack of sleep. An important proviso to this is probably that the whole transaction must take place in the relatively low region of the arousal continuum, that is along the ascending arm of the inverted-**U** function which Duffy (1934), Freeman (1948), and Hebb (1955) have postulated as a link between activation and performance; in other words the test situation must be uninteresting or *S* very sleepy.

The interaction of two agents which are known to influence respectively motivation and arousal thus supports the idea of a close relation between the two concepts and may even suggest a common site of action neurologically, the obvious choice at our present state of knowledge being the brain stem reticular formation.

The second question raised by the present study concerns the influence of familiarity with the task and with the experience of going without sleep. Over a period of 6 wk. the present *S*s were repeatedly exposed to the same task and the same loss of one night's sleep. To the present author's knowledge this has been done only once before (Lee & Kleitman, 1923) but only one *S* was used and the outcome was inconclusive. The present results show at moderate but acceptable levels of significance that the influence of this familiarity was to heighten the effect of lack of sleep on the task. Three explanations may be considered: the first is that losing one night's sleep a week for 6 wk. has a cumulative effect. This seems unlikely; present *S*s

had a whole week to recover from each night awake and the admittedly limited evidence we have suggests that one or two nights' normal sleep will suffice to remove the effects of a term of sleep-deprivation much longer than this. The second possibility is that after *Ss* have foregone a night's sleep on six separate occasions the experience loses its novelty; *Ss* may no longer feel challenged to make the extra effort which has often been held responsible for the failure of lack of sleep to affect performance (Edwards, 1941; Freeman, 1932; Robinson & Hermann, 1922). In the present experiment the decline of this stimulus towards the end of the 6-wk. program may have produced a more realistic and radical picture of the effect of lack of sleep on performance. Finally, having carried out the test twice a week for 6 wk., the test, rather than the experience of losing sleep, may have lost its novelty. It may be that one effect of lack of sleep is to increase the tendency for attention to be distracted from the main task by extraneous factors, especially when the task is one which does not compete well for attention. Berlyne (1950) has shown that decreased novelty can reduce a task's ability to compete in this way and this might explain why 6-wk. familiarity with the test increases the effect of lack of sleep upon it. Perhaps an important related point here is that the final state of decreased task novelty may also be a more realistic one in that many everyday tasks have reached at least a comparable state of familiarity. The lesson here may be that if we wish to assess the effect of lack of sleep and, indeed, that of other related stresses, for example noise and heat, on such everyday activities we may obtain more realistic answers from *Ss* who, as in the present experiment, have been made thoroughly familiar through practice with the stress.

Before leaving this matter of the interaction of time with lack of sleep let us consider what happens during the individual test. It is well known that the effect of lack of sleep on performance can become greater with time spent continuously on the test, and we have seen this happen in the present experiment. But as a similar effect has appeared over the more protracted period of 6-wk. intermittent testing we can now compare the patterns of these two interactions. Interesting similarities exist. A no-sleep impairment in corrects has become noticeable by the second 5-min. period of the test and by the second 2 wk. of the 6-wk. program. In gaps this does not appear until the third 5 min. of test and the third 2 wk. of the program. Also errors show a unique pattern of negative impairment passing to positive impairment along both the 30-min. and the 6-wk. time scale. This correspondence suggests that the nature of the fatigue implied by the interaction may be the same whether time spent on the test is continuous or intermittent. This conclusion is in accord with that of an earlier study (Wilkinson, 1959) which discovered no difference in no-sleep impairment when a continuous test of 25 min. was compared with a similar one interspersed with 30-sec. breaks. The influence of both these results should be towards lessening the emphasis placed by Williams, Lubin, and Goodnow (1959) on the importance of *uninterrupted* work in contributing to a no-sleep impairment. It is tempting to conclude that the main casual factor in the interaction of lack of sleep with time on task, whether massed or spaced, is decreasing novelty. Our confidence in this interpretation is strengthened if we can assume that decreased novelty implies reduced reward from the activity of attending to the test, for then the conclusion reflects in a different context (the interaction of time with lack of sleep) the main finding of this experiment, namely that lowered reward (reduced KR) increased the effect of lack of sleep.

Wide individual differences exist in the extent to which lack of sleep impairs performance, and in studies of the subject it has been commonplace to note this. In the present experiment the consistency of these differences when assessed independently over three successive 2-wk. periods has shown them not to be due to chance.

This knowledge should encourage the search for measures, perhaps of personality, physique, or neurophysiological makeup, which correlate with them.

Summary

The effect of one night's loss of sleep on a test of serial reaction was largely removed when the test was changed to allow greater feedback of knowledge of results. This interaction of lack of sleep with a known motivational factor suggests that the stress has one of its main effects in this area. To the extent that lack of sleep can be expected to lower arousal the result also supports suggestions of a close link between the concepts of arousal and motivation.

A basic 2-wk. program of testing was repeated three times on each *S* so that the effect of lack of sleep could be assessed independently in each of the three 2-wk. periods. The adverse effect on performance increased over the 6 wk.; three possible causes are considered: (*a*) a cumulative effect of loss of one night's sleep a week; (*b*) increasing familiarity with the stress; (*c*) increasing familiarity with the test. The two last-named explanations are preferred. It is suggested that the reduced novelty of both the stress and the test which prevailed during the last 2 wk. represents a more realistic situation; similar "acclimatisation" to the experimental situation might be of value in the examination of the effects of other stresses, for example, noise and alcohol, where factors of motivation and novelty may be important.

A further finding was that in the three identical and independent assessments of the effect of lack of sleep on the same *Ss*, it was generally the same men whose performance was most impaired. Real individual differences, often suspected, are now confirmed and should be examined more closely.

References

Armington, J. C., & Mitnick, L. L. Electroencephalogram and sleep deprivation. *J. appl. Physiol.*, 1959, **14**, 247–250.

Bartoshuk, A. K. Electromyographic gradients as indicants of motivation. *Canad. J. Psychol.*, 1955, **9**, 215–230.

Bartoshuk, A. K. EMG gradients and EEG amplitude during motivated listening. *Canad. J. Psychol.*, 1956, **10**, 156–164.

Berlyne, D. E. Novelty and curiosity as determinants of behaviour. *Brit. J. Psychol.*, 1950, **41**, 68–80.

Bjerner, B. Alpha depression and lowered pulse rate during delayed actions in a serial reaction test. *Acta physiol. Scand.*, 1949, **19**, Suppl. 65.

Duffy, E. Emotion: An example of the need for reorientation in psychology. *Psychol. Rev.*, 1934, **41**, 184–198.

Edwards, A. S. Effects of the loss of 100 hours of sleep. *Amer. J. Psychol.*, 1941, **54**, 80–91.

Elwell, J. L., & Grindley, G. C. The effect of knowledge of results on learning and performance. *Brit. J. Psychol.*, 1938, **29**, 39–53.

Freeman, G. L. Compensatory reinforcements of muscular tension subsequent to sleep loss. *J. exp. Psychol.*, 1932, **15**, 267–283.

Freeman, G. L. *The energetics of human behavior*. Ithaca: Cornell Univer. Press, 1948.

Hebb, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

Lee, M. A. M., & Kleitman, N. Studies on the physiology of sleep: II. Attempts to demonstrate functional changes in the nervous system during experimental insomnia. *Amer. J. Physiol.*, 1923, **57**, 141–151.

Leonard, J. A. 5 Choice serial reaction apparatus. *Med. Res. Council Appl. Psychol. Res. Unit Rep.*, 1959, No. 326/59.

Malmo, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–287.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

Pepler, R. D. Warmth and lack of sleep: Accuracy or activity reduced. *J. comp. physiol. Psychol.*, 1959, **52**, 446–450.

Robinson, E. S., & Hermann, S. O. Effects of loss of sleep. *J. exp. Psychol.*, 1922, **5**, 19–32.

Schnore, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *J. exp. Psychol.*, 1959, **58**, 117–128.

Siegel, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

Stennett, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61.

Surwillo, W. W. Psychological factors in muscle-action potentials: EMG gradients. *J. exp. Psychol.*, 1956, **52**, 263–272.

Wallerstein, H. An electromyographic study of attentive listening. *Canad. J. Psychol.*, 1954, **8**, 228–238.

WHITING, H. F., & ENGLISH, H. B. Fatigue tests and incentives. *J. exp. Psychol.*, 1925, **8**, 33–49.

WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid, 1949.

WILKINSON, R. T. The effects of lack of sleep on perception and skill. Unpublished doctoral dissertation, University of Cambridge, 1957.

WILKINSON, R. T. The effects of sleep loss on performance. *Med. Res. Council Appl. Psychol., Res. Unit Rep.*, 1958, No. 323/58.

WILKINSON, R. T. Rest pauses in a task affected by lack of sleep. *Ergonomics*, 1959, **2**, 373–380.

WILLIAMS, H. L., LUBIN, A., & GOODNOW, J. J. Impaired performance with acute sleep loss. *Psychol. Monogr.*, 1959, **73**(14, Whole No. 484).

(Received July 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 272-282

FIGURE COHERENCE IN THE KINETIC DEPTH EFFECT

BERT F. GREEN, JR.

Lincoln Laboratory,[1] *Massachusetts Institute of Technology*

When an observer views the two-dimensional (2-D) projection, e.g., shadow, of a moving three-dimensional (3-D) object, he usually perceives the shadow pattern as a form with depth. This has been called the Kinetic Depth Effect (KDE). Wallach and O'Connell (1953) concluded that an essential condition for the occurrence of the KDE seemed to be contours or lines that change their direction and their length simultaneously. Wallach, O'Connell, and Neisser (1953) showed that experience with an unfamiliar figure undergoing the KDE led to later perception of depth in the stationary shadow of the object. Gibson and Gibson (1957) studied the apparent slant of a 2-D surface formed by a set of regular or irregular forms. They obtained accurate judgments of slant when these forms underwent the continuous perspective transformations associated with plane rotation. Their experiments could be considered instances of the KDE but the Gibsons emphasize the importance of perspective for the perception of rigid motion, while perspective was apparently not an important determinant in the studies of Wallach et al.

In the KDE the relative motion of the various parts of a 3-D figure provides information not only about depth but also about the shape of the figure. The visual system in some way transforms the relative motions of the shadow's elements into the perception of a single rigid object, and this process appears to be a major factor in our normal perception of three-dimensional objects in the world around us. Johansson (1958) states the case as follows: "Mathematical relationships in the continuously changing energy distribution on the retina may be substituted for the classical static retinal picture, as the source of information from the external world. The substitution is viewed as an application of Gibson's gradient theory" (p. 3). It follows that a set of isolated dots or unconnected lines undergoing the appropriate changes can be perceived as a coherent rigid figure. White and Mueser (1960) studied the accuracy of the perception of arrangements of elements in a KDE setting. They found that all of their figures were perceived as rigid configurations in 3-D, but that reproducing the exact spatial relationships among the elements was a difficult task.

This paper reports a series of experiments designed to isolate the effect of relative movement from all other cues to depth and coherence. The experiments explore some of the conditions under which the KDE imparts perceived rigidity (coherence[2]) to elementary patterns of dots and lines. The *O*s viewed the changing 2-D projection associated with a rotating rigid 3-D configuration of dots or straight lines, and judged the apparent rigidity and coherence of the configurations, i.e., the extent to which the elements appeared to maintain their relative positions in the configuration. Independent variables were the num-

[1] Operated with support from the United States Army, Navy, and Air Force.

[2] Rigidity and coherence are treated as synonyms in this paper.

ber of elements in the configuration, the amount of constraint on the placement of elements, the type of rotation, the type of figure, and the speed of rotation.

General Procedure

The stimuli were 16-mm. motion pictures showing the 2-D projections of rotating 3-D figures and were made by an animation technique. Each frame of the film showed the figure's 2-D projection from a certain 3-D orientation. Successive frames showed projections from a succession of different orientations that corresponded with a particular rotation of the figure. A succession of 250 frames formed the film strip for a figure. Projected at a rate of 20 frames per sec., the strip gave a 12.5-sec. movie of a rotating figure. The films were produced frame-by-frame, by an automatic camera that photographed a display generated on a cathode-ray tube by a digital computer. (The M.I.T. Lincoln Laboratory's Memory Test Computer, which has since been dismantled, was used to generate the stimuli in Exp. I, II, and III. The Laboratory's I. B. M. 704 computer was used for Exp. IV, V, and VI.) The computer was programed to display the 2-D perspective projections of a set of isolated dots or straight lines whose 3-D coordinates were stored in the computer. After making the display, the computer actuated the camera, advancing the film by one frame. The computer then calculated new 3-D coordinates for all the points or lines according to specified rotation formulas and repeated the cycle, displaying the 2-D projection, advancing the film, and rotating the figure, until the specified number of frames had been photographed. The particular series of pictures to be made on any computer run was controlled by the initial parameters, which included the initial coordinates of the dots or lines, the type of rotation, and the parameters of the rotation formulas. The computer program, in FORTRAN language, is available from the author. It is appropriate for the IBM Type 704 and 709 computers with an on-line CRT display.

The regular figures that we used were symmetric about the origin, with the origin being the centroid of the figure. The random figures were samples from a population for which the origin was the centroid, on the average. Two types of rotation were used: spinning about a fixed axis through the origin, and tumbling about the origin. In the spinning rotation, the figures revolved about a fixed axis as the earth revolves about its polar axis. The axis itself was not shown—it was merely used to specify the direction of movement. (The fixed axis is the locus of points that do not move.) In addition to specifying the axis, the programer specified the angular velocity of the spin about the axis. Except in Exp. IV, where speed was varied, the speed was always 64°/sec for all figures. In the tumbling rotation, the origin of the 3-D coordinate system was the only point that remained fixed. From any one frame to the next the tumbling rotation amounted to a spin about an axis, but throughout the film strip, the orientation of the axis changed continuously. To tumble the figure about a fixed point (the origin of the coordinate system) the programer used Euler's formulas (Snyder & Sisam, 1941, p. 42) to specify the rates of change of three angles, representing components of rotation about each of the three coordinate axes in turn. The speed of the tumbling rotation, as measured by the spin about the moving axis, was not constant. Also the movement of the axis did not have a constant velocity. Nevertheless the average angular velocity could be controlled; it was equated to the constant velocity of the spinning rotations.

The 2-D projections included perspective, which was computed in terms of the ratio of nominal figure diameter to nominal viewing distance. In all our experiments this ratio was about $\frac{1}{18}$, which is so small that the stimuli would have been nearly the same had we used a parallel 2-D projection rather than a perspective projection. The *O* sat 9 ft. from the projected display, on which the figure diameter was about 1 ft., making the visual angle $6\frac{1}{2}°$. On the projected display, dots and lines were white on a dark ground. All dots had the same brightness and size, as did all lines.

Each configuration was displayed, rotating, for about 12.5 sec. The *O* was informed truthfully that he was actually viewing the 2-D projection of a rigid 3-D configuration, but he was instructed to rate the configuration as it appeared to him, rather than as he knew it to be. The *O* was told to use a 5-point subjective scale of rigidity or coherence, on which he was to give a rating of 1 if all the elements maintained their relative positions in the configuration throughout the exposure, and a rating of 5 if elements appeared to be moving independently. Intermediate ratings were to be given according to the proportion of coherent elements, and the relative amount of time that coherence was perceived. The

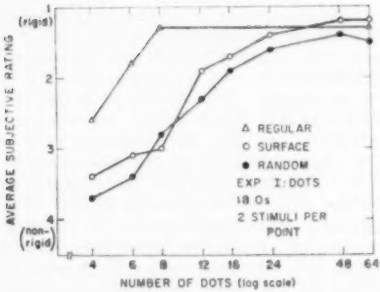

FIG. 1. Average ratings of rigidity in Exp. I, for three types of dot configurations. (In this and subsequent figures, the chance variability cannot easily be portrayed, because of the large individual differences. To assess the differences between any two figures, a difference in mean ratings of about ½ rating step is roughly equivalent to a significant preponderance of ratings of one figure over another. When comparing points each based on two figures, the equivalent mean difference is about ¼ rating step.)

Os had no difficulty making the ratings after five practice trials. All *Os* were members of the laboratory staff, and were familiar with psychophysical experiments. Each *O* was tested individually, and made his responses orally.

EXP. I: CONSTRAINT AND NUMBER OF DOTS

Procedure

The first experiment used configurations of dots under three conditions of constraint. In the *random* condition, the 3-D coordinates of the dots were chosen so as to keep the dots within a 3-D cubical confine, i.e., each coordinate was chosen at random from the interval $-c \leq x \leq c$, where 2c is the length of one side of the confine. In the *surface* condition, dots were placed on the surface of a hypothetical cube or double pyramid (the latter has six vertices, $[0, 0, \pm c]$, $[0, \pm c, 0]$ and $[\pm c, 0, 0]$). In the case of the cube this was accomplished by setting one coordinate to $\pm c$, and choosing the other two at random. A similar method was used for the double pyramid. Each surface had the same number of dots as far as possible. For both surface and random conditions, eight values of n, the number of dots, were used: 4, 6, 8, 12, 16, 24, 48, and 64. In the *regular* condition, dots were placed at the vertices of a regular tetrahedron ($n = 4$), a regular double pyramid ($n = 6$) a cube ($n = 8$), and in a $4 \times 4 \times 4$ regular cubical array ($n = 64$). The tumbling rotation was used.

Eighteen *Os* were used. Two stimulus films were made for each of the 20 conditions (the regular conditions were repeated, but new random samples were drawn for the random and surface conditions). The 40 film strips were arranged at random in two blocks of 20 and spliced into a single film. Nine *Os* viewed the film projected in the forward direction, the others saw the film projected in reverse—they not only received the reverse order of stimuli, but also saw the stimuli moving in the reverse direction.

In this and subsequent experiments, the data will be reported in terms of the average ratings of coherence, pooling *Os* and stimuli within conditions. An attempt was made to scale the stimuli by the method of successive intervals. The results were in very close agreement with the simple averages of ratings, except for stimuli that most *Os* rated "1". For these, the scale values were very erratic, being violently affected by two or three ratings other than "1." Because of this instability, the scaling results will not be reported.

Results

The average ratings of coherence are shown in Fig. 1. Clearly, the regular configurations were judged to be more coherent than the surface and random configurations, for small n, while for 64 dots, all conditions yield the same high degree of perceived coherence. There is also a small but consistent difference between surface and random configurations, leading to the conclusion that the amount of constraint or regularity affects perceived coherence. The strong effect of the number of elements, n, is also clear from Fig. 1; the more elements, the more coherent a configuration appears.

An analysis of variance was made for the data from the random and surface constraints. The *Os* were divided into two groups, according to the order of presentation of stimuli,

TABLE 1

ANALYSIS OF VARIANCE OF RATINGS OF COHERENCE IN EXP. I

Source	*MS*	*df*	Error Term	Error *df*	*P*	Variance Components
Number of dots (*n*)	6.4046	7	.3100[a]	12.2	.001	.7618
Groups (G)	2.9714	1	.1954	7	.01	.0868
Constraint (C)	.7855	1	.1415[b]	23	.05	.0201
NC	.0640	7	.1753	16	—	—
NG	.1954	7	.0607	16	.05	.0336
CG	.0123	1	.0607	16	—	—
NCG	.0955	7	.0607	16	—	—
NC(S)[c]	.1753	16	.0607	16	.05	.0573
NCG(S)	.0607	16	—			.0607

Note.—Procedures and notation follow Green and Tukey (1960).
[a] NG + NC(S) − NCG(S).
[b] NC(S) and NC pooled.
[c] Stimuli nested in NC.

giving an analysis of 2 constraints, 8 numbers of elements, 2 groups, and 2 stimuli nested in the combination of number of constraint. The analysis, shown in Table 1, indicated that both *n* and Groups have strong effects. The Group effect includes individual differences as well as presentation differences, and we suppose that the former predominates. The effect of constraint is barely significant at the .05 level. The significant *n* × Group interaction reflects the fact that the group differences, which in this analysis are surrogates of individual differences, are more apparent in the low values of *n*. For large *n*, most of the responses are "1." The analysis shows a large effect for *n*, and a strong effect of Group. The significant mean square for stimuli within *n* × Constraint indicates a small effect of particular stimuli.

EXP. II: LINES AND ROTATIONS

Procedure

The configurations in Exp. II were random line segments, the endpoints being specified in the same way as the dots in the random condition of Exp. I. The lines were either unconnected and independent, or connected, one to the next and the last to the first, to form a single closed 3-D curve. No 3-D curve lay in a plane—all had three dimensions. Two rotations were used: spinning about a vertical axis and tumbling. Five values of *n*, the number of lines, were used: 4, 6, 8, 12, and 16. Thus there were 20 conditions: 2 rotations × 2 types of figures × 5 values of *n*. Two stimulus movies were prepared for each condition, and were arranged in random order within each of two blocks of 20. Unfortunately the stimuli for the condition of 16 unconnected spinning lines were faulty and could not be replaced. Sixteen *O*s were used; most had served in Exp. I.

Results

The average ratings are shown in Fig. 2, averaged over *O*s and stimuli within conditions. Clearly the connected figures are seen as more coherent than the unconnected, and the spinning figures are seen as more

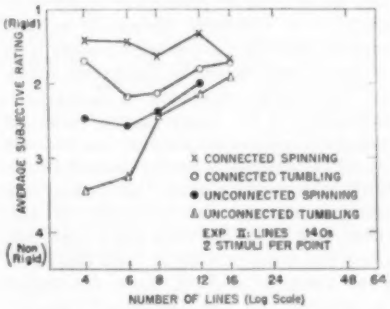

FIG. 2. Average ratings of rigidity for straight line figures in Exp. II.

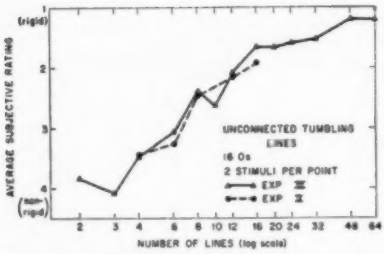

FIG. 3. Average ratings of rigidity for unconnected tumbling figures in Exp. III.

coherent than the tumbling. The statistical significance of these effects is established by a sign test. The effect of n is negligible for the connected lines, minor for the connected spinning figures, and important only for the unconnected tumbling figures.

The favorable effect of the spinning rotations is interesting. Rotation about a vertical axis is the most common kind of transformation seen when one walks about in the world. It is the type of rotation used exclusively by Wallach and O'Connell (1953). It is also a simpler rotation than tumbling, since every element moves at a constant angular velocity, and each element moves in a single plane, perpendicular to the axis of rotation. The questions of simplicity and familiarity are studied further in Exp. V and VI below.

The favorable effect of connected lines may be related to reversals. Since perspective is virtually absent from the stimuli, and since there are no brightness or size cues, depth is imparted solely by the KDE, so there is front-back ambiguity. In many cases, lack of coherence occurs because some, but not all, of the elements have reversed for *O*. With the unconnected lines, any one line can reverse independently of any other, but when the lines are connected, a reversal must involve at least two lines. Thus there is slightly more constraint on reversal in the connected figures, and this may account in part for the difference between the two types of figures. If the connected lines are interpreted as being more constrained than the unconnected, then the results of Exp. II are consistent with the effect of constraint found in Exp. I.

EXP. III: NUMBER OF UNCONNECTED LINES

Procedure

The third experiment extended the range of n in Exp. II for the unconnected tumbling figures, and also checked on the reliability of ratings. The stimuli for Exp. III were those from the unconnected tumbling condition of Exp. II, plus similar stimuli for additional values of n to cover the range of n from 2 to 64. In all there were two stimuli at each of 13 values of n, arranged at random in two blocks of 13 stimuli. The 16 *O*s used in Exp. II also served as *O*s in this experiment.

Results

The average ratings of coherence are shown in Fig. 3, together with the comparable curve from Fig. 2. It is clear that the ratings are adequately reliable, since the curves are about the same. Further, the effect of n for these figures is regular. Roughly, the average rating is linearly related to log n, at least up to 24 lines. Beyond 24, the average ratings decrease very little. The curve is very similar to the curve for the random dot figures in Fig. 1.

EXP. IV: SPEED OF ROTATION

Procedure

The fourth experiment was designed to study the effect of speed of rotation on judgment of coherence. Very slow and very fast speeds were expected to reduce the effectiveness of the KDE. Five speeds were used: 16, 32, 64, 128, and 256° per sec. The middle value, 64°/sec, was the speed used in all previous experiments. Five types of figures were used: 4, 16, or 64 random dots; 4 or 16 random unconnected lines. Only the tumbling type of rotation was used. Two versions of each speed × figure combination were prepared, yielding 50 stimuli. Each of 14 *O*s, most of whom had served in the previous

experiments, saw the 50 stimuli in the same random order.

Results

The results are shown in Fig. 4. Clearly there is very little effect of speed over the range in our experiment. There is a hint of slightly less coherence at the slowest speed, but the other four speeds, covering an eightfold range, show no differential effects on coherence.

The range of speeds covered in this experiment nearly spans the range of possible speeds in our experimental procedure. Another factor of two faster would give so much change from one frame to the next (25°) that apparent movement would probably be destroyed. Another factor of two slower would result in a total excursion of only about 100° in the viewing time for each figure; no point in the figure could move much more than half way across the display, so that to get meaningful judgments of coherence, the viewing time would have to be increased. White and Mueser (1960) found a similar effect of speed on accuracy of perception in their KDE setting. They obtained differences among speeds of 2, 8, and 32 rpm in a spinning rotation, which correspond with 12°/sec, 48°/sec, and 192°/sec. They found reduced accuracy with a slower speed of 1 rpm (6°/sec), but they used an exposure time of 30 sec., and their results suggest that a longer viewing time would have allowed greater accuracy. Slow speed by itself may not detract from the KDE.

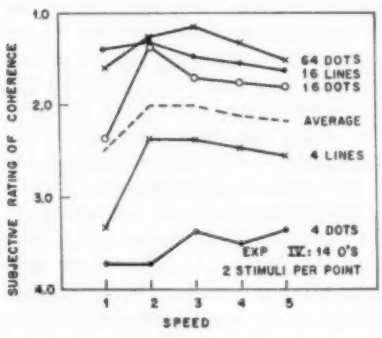

FIG. 4. Average ratings of rigidity for figures tumbling at five speeds in Exp. IV.

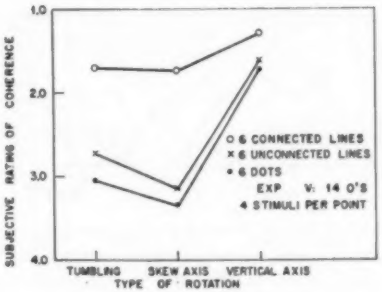

FIG. 5. Average ratings of rigidity for figures under three types of rotation in Exp. V.

EXP. V: SKEW SPINS

Procedure

A large difference was obtained in Exp. II between tumbling and vertically spinning. Experiment V was intended as a demonstration that spinning about a nonvertical axis would yield coherence judgments that were somewhere between those obtained with a vertical spin and with tumbling. The notion was that the judgments of coherence were related to the complexity of the rotation.

Three types of rotation were used: tumbling, spinning about a vertical axis, and spinning about a skew axis that was at an angle with all three coordinate axes. Two different skew axes were used. The cosines of the angles made by one skew axis with the horizontal axis in the display plane, the vertical axis, and the axis perpendicular to the display plane, were in the ratios 1:2:3, respectively. The other skew axis had cosines in the ratios 1:1:1. Three types of figures were used: six random dots, six unconnected random lines, and six connected random lines. Four versions of each figure type were made for the vertically spinning and the tumbling rotations, and two versions of each were made for each skew axis, yielding 36 stimuli. The 14 *O*s who had served in Exp. 4 were enlisted for Exp. 5.

Results

The results are shown in Fig. 5. Since the results for the two skew

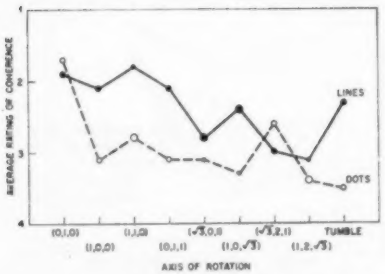

FIG. 6. Average ratings of coherence for various rotations in Exp. VI. (Small points represent 16 *O*s rating one stimulus; medium points, 16 *O*s rating 2 stimuli; large points, 16 *O*s rating 3 stimuli.)

axes did not differ more than differences between stimuli for a given axis, the results for the two skew axes were averaged. The results for the skew axes are in no sense intermediate between tumbling and vertically spinning. If anything, tumbling is intermediate. Clearly the effect of axis of rotation is more complicated than we had supposed. Another experiment was designed to study a variety of axes of rotation, in an attempt to resolve the complication presented by Exp. V.

EXP. VI: AXIS OF ROTATION [3]

Procedure

Nine patterns of rotation were used: the tumbling rotation and eight different fixed axes with the spinning rotation. The position of an axis is best specified by its direction numbers, which are numbers proportional to the cosines of the angles made by the axis with the three coordinate vectors: the horizontal vector in the display plane, the vertical vector, and the vector perpendicular to the display plane, respectively (Snyder & Sisam, 1941). The eight fixed axes of rotation used in Exp. VI can be divided into three groups. Group 1 includes axes in the plane of the display: (0, 1, 0), the vertical axis; (1, 0, 0), the horizontal axis in the display plane; and (1, 1, 0), an axis at an angle of 45° with both the horizontal and vertical vectors in the display plane. Group 2 includes axes not in the display plane but perpendicular to one reference axis: (0, 1, 1), an axis tilted forward 45° from the vertical and 45° up from the axis perpendicular to the display plane; ($\sqrt{3}$, 0, 1) and (1, 0, $\sqrt{3}$), axes 30° and 60° back from the display plane, respectively, in a plane perpendicular to the vertical reference axis. Group 3 contains two skew axes that made acute angles with all three reference vectors: ($\sqrt{3}$, 2, 1) and (1, 2, $\sqrt{3}$), axes 30° and 60° forward from the display plane, respectively, and both at 45° from the vertical reference vector.

Two types of figures were used; six random dots and six random unconnected lines. Two replications of each figure type were intended to be made for each type of rotation, yielding 36 conditions. After the experiment had been completed, it was discovered that for axes (1, 1, 0) and (1, 2, $\sqrt{3}$) there were three dot figures and one line figure, while for axes ($\sqrt{3}$, 0, 1) and (1, 0, $\sqrt{3}$) there were three line figures and one dot figure, instead of two each. Sixteen *O*s, most of whom had served in the previous experiments, rated the displays for coherence in the usual way.

After the ratings had been obtained from all *O*s, 13 *O*s were tested again. This time *O* was asked to indicate the direction of the axis of rotation by setting a rod in a matching position. The *O* was instructed to make a setting, even if there did not seem to be a single fixed axis of rotation.

Results

The results of the coherence ratings are shown in Fig. 6. For the dot figures, all fixed axes but the familiar vertical lead to about the same level of coherence, and the tumbling dots are also at this level. Significantly more coherence is obtained by spinning about the vertical axis. For the line figures, the vertical axis also leads to relatively high coherence, but so do the other axes in the display plane, the tilted vertical axis, and the tumbling rotation. The other skew axes impart less coherence. The coherence ratings of the line figures support the hypothesis that the more complex axis locations are associated with less

[3] The author is indebted to Alice Wolf for running part of this experiment.

coherence. The three groups of axes defined above were formed on a priori grounds to represent three levels of complexity. When the ratings for each *O* are averaged across all line figures within each group of axes, we find that for 11 of the 16 *O*s the group averages are in the order 1, 2, 3, a preponderance significant at the .0001 level. Very little support for the hypothesis can be obtained from the ratings of the dot figures, since only the vertical axis stands out.

The results of the rod settings are shown in Fig. 7, where the proportion of setting within 15° and within 30° of the true axis are plotted for each axis. In order to plot the settings from the tumbling figures on the same figure, the median setting for the tumbling displays was taken as the "true" axis (the median was determined separately for dots and lines but the judgments from the two replications of each figure were pooled). The settings for the axes not in the display plane provide evidence for the reversals seen by *O*s. All of the true axes were in front of the display plane (the end of the axis above the origin of the display is used for this reference). Actually 35% of the settings had the axis in front, 50% in back, and 15% on the display plane. Thus the figures were seen at least as often with front and back reversed as in the programed orientation. For this reason, reversals were not counted as errors. There were two equally correct settings for all axes not in the display plane. On this account, the chance level for Group 1 differs from that of Groups 2 and 3. Chance levels are computed on the basis of a uniform distribution of settings over the surface of a unit sphere, and are shown in Fig. 7.

Figure 7 shows that the settings for the tumbling figures were very

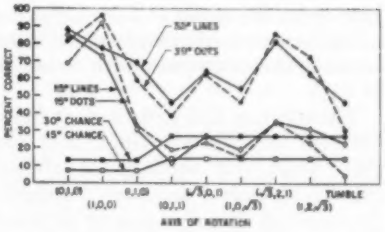

FIG. 7. Proportion of axis settings within specified angle of the true axis, in Exp. VI. (Two criteria of correctness are shown; 15° and 30°. Computed chance levels are shown for both criteria.)

inaccurate, in terms of the arbitrary "true" axes, indicating that the settings were quite variable. For the fixed axes, the stringent criterion of 15° error separates the vertical and horizontal axes from all the rest, and shows no differences among the rest. It also shows that setting the axis was a difficult task not accurately done by the *O*s. The sparse figures (only six elements in each), the short viewing time, and the tricky response device all contributed to the difficulty. Also, one or two *O*s occasionally set the rod at right angles to the true axis, apparently matching the direction of movement rather than the axis about which the movement occurred.

Measured by the liberal 30° criterion, the rod settings showed significant differences among the fixed axes, but no clear pattern emerges. Considering the three groups of axes described above, Groups 1 and 3 appear to have been matched with about the same accuracy but because of the differences in chance level, the settings for Group 1 would seem to be actually better. The axes in Group 2 were matched less accurately than the others. Individual differences in accuracy are huge. The best *O* had 97% of his settings within 30° of the

true axis; the worst *O* only 19%; the other *O*s were fairly evenly distributed between these extremes, with a median of 78%.

The dot and line figures elicit quite different coherence ratings, and yet they yield rod settings of almost exactly the same accuracy. Indeed, *O*s reported that the axis of rotation was apparent in the dot displays even when the figures did not cohere. That is, even if some dots "broke away" perceptually from the figure, the direction in which the majority of the dots were rotating was still clear. In the case of dot figures rotating about axes in the display plane, about half the *O*s reported sometimes seeing the dots moving in 2-D, sliding back and forth on their straight paths (with sinusoidal velocity) rather than rotating in a 3-D circular path (with constant velocity). Nevertheless the proper setting for the rod was obvious.

There are other indications that matching the axis was psychologically quite different from judging coherence. The r between the average coherence rating of a figure and the number of errors made in matching its axis of rotation was only 0.16. The rank correlation between the number of ratings of "1" given by an individual and the accuracy of his settings was −.45.

DISCUSSION

It is abundantly clear that the KDE alone is a very powerful factor in depth perception. Many of our *O*s saw every figure in 3-D; some *O*s reported occasionally seeing figures in 2-D, but the occasions were rare and fleeting, even with the meager figures that were used. The tendency to see depth in the 2-D projections of rotating 3-D figures is overwhelming. There is also a strong tendency for points or line segments undergoing the same 3-D rotation to be seen as parts of a rigid, coherent figure. Even randomly placed, unconnected points and lines form a rigid figure if they undergo the same 3-D rotation, while connecting the elements, or otherwise constraining them enhances the tendency to perceive a rigid figure.

The KDE is powerful, but it is not a *sufficient* condition for perceiving complete coherence. The majority of our figures were perceived as 3-D but nonrigid. The "1" rating was used less than half the time, and the "5" rating was seldom used, indicating that *O*s judged most figures to be partially coherent. Matching the axis of rotation in Exp. VI was not done very accurately, despite the omnipresence of the KDE. White and Mueser (1960) found that the spatial arrangements of elements were not accurately perceived even though the configuration was perceived as rigid and three-dimensional. The KDE seems to be more immediate than perceived rigidity, and both seem more immediate than accuracy of perception of unfamiliar random figures.

The role of perspective in the KDE is not determined. Gibson and his coworkers (1957, 1959) emphasize the importance of changes in perspective transformations. In our experiments, on the other hand, perspective was essentially irrelevant. The *O*s reported that reversals were common, and the axis-matching data of Exp. VI show that there was no agreement about which was front and which was back. (In retrospect, it would have been better to use parallel projection in our experiments, to make our statements about perspective unequivocal.) The truth must be that perspective is effective when it has a chance to be; perspective is very important when it is essentially the only clue to depth, but it is no more an essential feature of depth perception than is any other single factor. Braunstein (1960)[4] has data that confirm this supposition.

Some of the characteristics of figures that influence perceived coherence have

[4] M. Braunstein, personal communication, 1960.

been established. The number of elements is of prime importance, as are connections and constraint among elements. From *O*s' reports we know that isolated elements have a tendency to "break away" from the main body of the figure, and that lines that cross each other seem to enhance coherence. But our experiments did not include curved lines, solid figures, or familiar forms. Much remains to be learned about the properties of figures that are related to the KDE.

There is evidence of consistent individual differences across experiments. Twelve persons served as *O*s in all six experiments. For each experiment these *O*s were ordered according to the number of stimuli given a rating of 1. The similarity of the six resulting rankings of the 12 *O*s was evaluated by Kendall's coefficient of concordance, *W*. We found that $W = .60$, which is not only significantly different from zero at well beyond the .001 level, but also indicates very considerable consistency. The consistency of individual differences is all the more remarkable because the experiments were carried out at various times during a 3-yr. period. In each experiment, the percentage of "1" ratings given by the various *O*s varies over a wide range; the average range is 53 percentage points. How much of such large, consistent differences is due to differences in criteria for the "1" response, and how much to differences in perception cannot be determined from our data.

The subjective rating procedure served its purpose well. The ratings were reliable, and gave adequate measures of the major effects studied in these experiments. However, the distributions of ratings obtained in various conditions overlapped considerably, and we feel that a more sensitive measure is needed to investigate some of the finer points of the KDE. Also, individual differences in perception are confounded with individual differences in subjective scale values. To extract the individual differences which are known to be large, one must take enough data to establish firmly each *O*'s subjective scale, or else one must use an objective procedure.

How accurate is depth perception in the KDE? We may assume that when *O* saw a figure as rigid, he saw either a true 3-D figure or a complete front-back reversal of the true figure since there was only one rigid 3-D figure that would produce the 2-D display. But we have no other evidence that *O* saw the true figures, and indeed, the inaccuracy in the rod settings of Exp. VI suggests the possibility that depth is not perceived accurately. Accuracy of depth perception probably depends on the number of elements and the coherence of the figure, but the nature of this dependence is unknown.

There are many other unsolved problems. Can *O* discriminate between a rigid and a nonrigid 3-D figure, when perceived in the KDE situation? What happens with only one or two points or lines, i.e., what is the minimal stimulus for the KDE? Why are the more complex axes of rotation associated with less coherence? When is perspective important? Must all of the elements of the figure be continuously in view, as they were in these experiments? (Casual observation, plus everyday experience, strongly suggest that the answer is "No.") A wide field for research exists.

Finally, the role of the digital computer in this research deserves comment. Developing the computer techniques and camera techniques were difficult and time consuming, but without the computer, we would not have done the experiments at all. In our present setup, a film strip of 250 frames showing one figure rotating for 12.5 sec. can be produced in about 2 min., and there is complete assurance that all the elements in the figure will be of uniform size and brightness. Of course, KDE experiments can be done without computers, but the computer provides great flexibility in the kinds of experiments one can envisage and perform.

Summary

Six experiments examined the extent to which the kinetic depth effect produces per-

ceived coherence and rigidity of random figures. Subjectively rated coherence was greater with (*a*) more elements in the figure, (*b*) connections and constraints among the elements, and (*c*) less complex axes of rotation. The tumbling rotation was shown to be intermediate between simple and complex axes of spinning rotations. Speed of rotation was shown to have almost no effect. Individual differences were large. A high speed digital computer was essential for producing the stimuli for the experiments.

REFERENCES

GIBSON, J. J., & GIBSON, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54**, 129–138.

GIBSON, E. J., GIBSON, J. J., SMITH, O. W., & FLOCK, H. Motion parallax as a determinant of perceived depth. *J. exp. Psychol.*, 1959, **58**, 40–51.

JOHANSSON, G. Rigidity, stability, and motion in perceptual space. *U. Uppsala Psychol. Lab. Rep.*, 1958, No. 3.

SNYDER, V., & SISAM, C. H. *Analytic geometry of space.* New York: Holt, 1941.

WALLACH, H., & O'CONNELL, D. N. The kinetic depth effect. *J. exp. Psychol.*, 1953, **45**, 205–217.

WALLACH, H., O'CONNELL, D. N., & NEISSER, U. The memory effect of visual perception of three-dimensional form. *J. exp. Psychol.*, 1953, **45**, 360–368.

WHITE, B. W., & MUESER, G. E. Accuracy in reconstructing the arrangement of elements generating kinetic depth displays. *J. exp. Psychol.*, 1960, **60**, 1–11.

(Received July 9, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 283–290

CONCEPT IDENTIFICATION USING SIMULTANEOUS AUDITORY AND VISUAL SIGNALS[1]

DANIEL S. LORDAHL[2]

University of Wisconsin

Among the variables affecting performance in concept identification tasks, the amount of irrelevant information in the stimuli has been most often studied. Archer, Bourne, and Brown (1955), Bourne (1957), Bourne and Pendleton (1958), and Brown and Archer (1956), have all varied the amount of irrelevant information in concept tasks. These studies differed with respect to other independent variables studied, but each showed that increases in irrelevant information caused significant increases in the difficulty of the task.

Bourne (1957), Bourne and Pendleton (1958), and Brown and Archer (1956) all found amount of irrelevant information (in bits) in problems of this type increases the difficulty of the task in a linear manner within the limits studied. Moreover, this effect holds over a number of minor variations in procedure. There were, however, a number of parameters in this set of experiments which, if varied, might upset the generality of the effect of irrelevant information. For one thing, each of the above studies used visual stimuli exclusively. A change of sense modality probably would not have changed the relationship between performance in concept identification and irrelevant information, but there is no assurance of this. Even less can be presumed about the effect of presenting information through two senses simultaneously. The present study investigates just this situation: the simultaneous presentation of information through the visual and auditory senses.

Some recent theoretical approaches have been made to the concept situation in terms of statistical learning theory (Bourne & Pendleton, 1958; Bourne & Restle, 1959; Walker, 1958). Although the present study was designed to answer the empirical problem indicated above, it also has implications for the mathematical theories of concept identification.

METHOD

Subjects.—The *Ss* were volunteers drawn from the elementary course and certain intermediate courses in psychology at the University of Wisconsin during the spring and summer of 1959. The data from 96 *Ss* are included in this paper. The data from four groups of 2 *Ss* were dropped on account of procedural or apparatus difficulties.

Design.—The major variables under consideration were amount of irrelevant information presented visually and amount of irrelevant information presented auditorily. Under all conditions, one stimulus dimension from *each* mode of presentation was relevant to the solution of the problem presented to *Ss*. Since only two levels of each dimension were used, in all conditions 1 visual and 1 auditory bit of information had to be taken into account by *Ss* in order to solve the problem.

Three different problems were used, i.e., three different pairs of relevant stimulus

[1] This paper is based on a dissertation submitted to the Graduate College of the University of Wisconsin in partial fulfillment of the requirements for the PhD degree. The writer is indebted to E. James Archer, who directed the study, for his aid and advice throughout the course of the experiment. The research was supported, in part, by funds from the Wisconsin Alumni Research Foundation.

[2] Now at Washington University, St. Louis, Missouri.

dimensions (one visual and one auditory). It should be noted that visual and auditory stimuli were presented simultaneously.

The design involved four amounts of irrelevant visual information, four amounts of irrelevant auditory information (0, 1, 2, or 3 bits in each case), and three problems. Two *Ss* served in each of the 48 conditions, with each *S* receiving 192 trials.

The *Ss* were required to make one of two lever-pressing responses for each of the 192 stimuli presented. All stimuli contained one of four possible combinations of the two bilevel relevant dimensions. One level from each relevant dimension was randomly designated as a *positive level*. When both positive levels were present in a given stimulus, that stimulus will be called a *positive instance* of the concept. All other stimuli will be called *negative instances* of the concept. When a positive instance of the concept was presented, the correct response for *S* was to press the positive lever, marked with a plus sign. The correct response for all negative instances was to depress the unmarked lever. These two responses, whether correct or incorrect, will be called *positive* and *negative lever responses*, respectively.

Apparatus.—The 2 *Ss* in each cell served at the same time and were seated at a long table with an opaque panel separating them. Each *S* was supplied with a set of earphones (crystal headphones, Model BA-200B; Clevite Electronic Components, Cleveland, Ohio) through which the auditory stimuli were presented. On the table, directly in front of each *S*, a gray box (6×3×3 in.) was fixed. This box had two black-handled lever-action switches (3-in. separation) extending toward *S* from the front side, and above one of the switches there was a large black plus sign on top of the box. To the right of each switchbox was a second box approximately 10 in. high which had two 4 × 4 in. milk glass screens (one above the other) in the side facing *S*. These screens provided *S* with information about the appropriateness of his lever-pressing response; the top glass was illuminated from within the box with a white light for a correct response and the bottom glass was illuminated with a red light for an incorrect response.

Ten feet in front of the 2 *Ss* was a gray partition separating *Ss* from *E* and his equipment. In the partition in front of the 2 seated *Ss* and about 1 ft. above eye level was a rectangular white translucent paper screen (12 × 10 in.) which was illuminated from behind by a strip film projector to provide the visual stimuli. The visual portion of the stimulus materials was photographed on 16 mm. film strips and projected on the screen from behind the partition.

The auditory signals were generated by an apparatus consisting basically of two audio oscillators and a white noise generator. Relay switching circuits varied the audio signal on five independent bileveled dimensions. The combinations of these dimensions in their different levels allowed for the possible presentation of 32 distinctly different auditory signals. The selection and presentation of the auditory signals was controlled by punched programs through Western Union Tape Transmitters.

The *Ss*' responses were recorded graphically with an Esterline-Angus 20-channel operations recorder and, in addition, errors were automatically totaled for each *S* on counters.

The information lights in front of *S* were lighted immediately upon his lever-pressing response and remained on until the next stimulus was presented, 2 sec. following the response of the slower *S*. Also present at the moment of stimulus change was a loud relay report which was audible to *Ss*. This sound was constant in pattern and intensity and as such gave no information to *Ss* other than that a new stimulus was present. The change from one stimulus to the next was automatically controlled so that the onsets of the visual and auditory portions were essentially simultaneous.

Task and instructions.—The *Ss* were told to place each combination of visual and auditory stimuli (called the "total signal") into one of two classes or categories by depressing one of the two levers on the boxes in front of them. It was pointed out that all of the total signals which belonged in the "plus class" (designated by the lever with a plus sign above it) had some aspect or aspects in common and that part of their task was to figure out which aspects of the total signal would allow them to correctly assign a signal to the plus class. The operation of the levers and the corresponding meaning of the information lights were explained to *Ss* along with the restriction that they choose one and only one lever to press for each signal.

The objective presented to *S* was to assign as many signals as possible to the correct class. They were told that by finding the solution to the problem (the relevant aspects of the signal) they could greatly increase their accuracy.

During a 1-min. rest in the middle of the session, *Ss* were told to relax but to not

discuss the problem. Before resuming the problem they were told that the problem was the same and that they should continue with the same objectives.

At the end of the session, each *S* was asked to write down the solution of the problem or any partial solution he had attained.

Stimuli and problems.—The visual stimuli, consisting of linear arrangements of a number of geometric forms, were off-white in color in contrast to the darker background of the paper screen. All visual figures contained a center fixation cross (+). The six dimensions of variation used and both levels of each were: (*a*) Size—large or small; (*b*) Shape—circles or ellipses; (*c*) Number of figures—4 or 3; (*d*) Orientation (linear direction of figures)—horizontal or vertical; (*e*) Location (on screen)—right or left; (*f*) Center spot—below fixation cross or above fixation cross.

The auditory stimuli varied along five dimensions, with two levels of each dimension. Easily discriminated levels for these five dimensions were determined in a prior psychophysical study: (*a*) Pitch of a pure tone—300 or 1800 cps; (*b*) Intensity of the tone—35 or 59 db. above threshold for the 300-cycle tone or 54 or 84 db. above threshold for the 1800-cycle tone; (*c*) Presence or absence of two short "blips" of white noise; (*e*) Steadiness of the pure tone or its regular interruption; (*f*) Earphone to which the signal was delivered— left or right.

When the variation of a dimension was not called for in a given treatment, the dimension was held constant at the level mentioned first in the above descriptions.

The temporal relationships between the noise and the tone interruption are illustrated in Fig. 1. It should be noted that the signal was recurrent, and could have started at any point on a given trial.

Three problems (A, B, and C) were used,

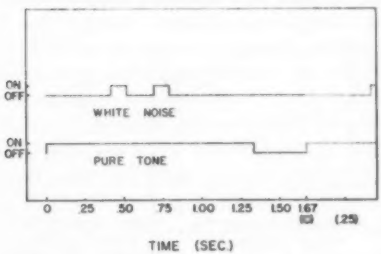

FIG. 1. Temporal relationship between the tone and noise portions of the auditory signal.

TABLE 1

RELEVANT AND IRRELEVANT DIMENSIONS USED IN EACH PROBLEM

Problems	Relevant (Positive Level)		Irrelevant					
	Visual	Auditory	Visual			Auditory		
			1	2	3	1	2	3
A	F (circle)	E (right)	L	O	C	I	W	R
B	S (large)	R (steady)	L	O	C	I	W	E
C	N (four)	P (high)	L	O	C	I	W	E

Note.—The letters in the table refer to the following dimensions: Center spot (C), earphone (E), intensity (I), form (F), location (L), number (N), orientation (O), pitch (P), interruption (R), size (S), and white noise (W).

selected by randomly pairing three auditory dimensions with three visual dimensions and obtaining three pairs of relevant dimensions. The selection of the particular levels of the relevant dimensions used to define a positive instance of the concept was at random.

Table 1 presents the relevant dimensions and the positive instance for each problem as well as the irrelevant dimensions for each. In reading Table 1, the numbered columns under "Irrelevant" indicate the order of adding irrelevant dimensions. Because of a limited selection of auditory stimuli, the third irrelevant auditory dimension in Problem A was different from Problems B and C.

Stimuli were arranged in a series of 96 which was repeated during the second half of the task. The positive instances of the concept, in the same locations for all problems, were placed in the series by random sampling without replacement, as were all other combinations of the stimuli. This allowed for complete balancing of each level for all dimensions as well as for the almost complete balancing of the number of times each level of a given dimension occured with particular levels of all other dimensions. The balancing of levels was done between each mode of presentation (auditory and visual) as well as within each mode. It should be noted that 25% of the stimuli were positive instances of the concept while 75% were negative instances.

RESULTS

Errors.—As the amount of irrelevant visual information was increased

TABLE 2

ANALYSIS OF VARIANCE OF TOTAL ERRORS

Source	*df*	*MS*	*F*
Problems (P)	2	86.391	
Irrel. auditory (A)	(3)	(277.394)	1.37
Linear	1	712.426	3.52
Quadratic	1	.047	
Cubic	1	119.709	
Irrel. visual (V)	(3)	(2772.852)	13.69**
Linear	1	7975.301	39.37**
Quadratic	1	282.755	1.40
Cubic	1	60.501	
P × A	6	139.342	
P × V	6	147.884	
V × A	9	279.079	1.38
P × V × A	18	229.548	1.33
Ss in cells[a]	48	202.589	4.77**
Halves of task (H)	1	9310.255	219.36**
H × P	2	87.099	2.05
H × A	3	8.852	
H × V	3	120.089	2.83*
H × P × A	6	30.175	
H × P × V	6	33.745	
H × V × A	9	38.982	
H × P × V × A	18	14.256	
H × *Ss* in cells[b]	48	42.443	
Total	191		

[a] Error estimate for terms above "*Ss* in cells."
[b] Error estimate for terms containing "Halves."
* $P < .05$.
** $P < .001$.

from 0 to 3 bits, the mean percentage errors increased as follows: 19.3, 21.8, 29.4, and 36.9. An extended trend analysis of variance (Grant, 1956) of the error scores (Table 2) indicated that this increase in errors with increased irrelevant visual information was very significant, and that a significant linear component was the only reliable source of variation. (A Bartlett's test for homogeniety of variance among the conditions yielded a chi square of 58.3 which, with 47 *df* does not force the rejection of the hypothesis of homogeneous variances.) The variations in errors with different amounts of irrelevant auditory information was not significant in this analysis.

The average percentage errors for each *S* was 34.0 for the first half of the task and 18.7 for the second half. This difference was highly significant (Table 2). In addition, there was a significant interaction between the amount of irrelevant visual information and the halves of training. This relationship is shown graphically on the left side of Fig. 2. It can be seen that the decrease in errors from the first to second half of the task became less as the amount of irrelevant visual information in the problem increased.

In addition to the total errors, certain types of errors were of special interest. When the relevant dimension of the visual signal on a given trial was at the positive level, and at the same time the relevant auditory dimension was at the negative level, *S* could have incorrectly selected the positive lever response. This type of error was defined as a Type A error. When the situation was reversed and the relevant visual signal was at the negative level while the relevant auditory signal was positive, the error was called a Type V error. Since the negative levels of each relevant dimension were consistently related to a negative instance of the concept, the Type A errors reflect a tendency of *S* to ignore the relevant auditory information.

These two types of errors were tabulated for each *S* for each half of the task. The *Ss* averaged 14.23 Type A errors and only 9.97 Type V errors during the whole task. The differences between the two types of errors were not related to the number of bits of irrelevant information.

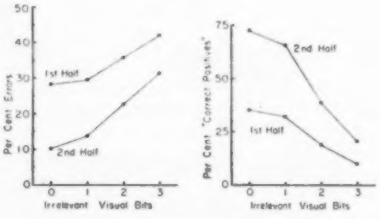

FIG. 2. Performance on first and second halves of the task as a function of the amount of irrelevant visual information.

The difference between the frequencies of the two types of errors was quite significant ($F = 15.96$, 1/80 *df*, $P < .001$) as was the interaction of *Ss* by types of errors ($F = 3.39$, 80/80 *df*, $P < .001$). Of interest also is the fact that the interaction of types by halves of the task was not significant ($F = .49$, 1/80 *df*), which would indicate that practice in this task had little if any influence on the predominance of the Type A errors.

Correct positive instances.—Since 75% of the stimuli presented in the task were negative instances, *Ss* could substantially reduce their errors by selecting the negative lever response more frequently than the positive response. Moreover, the use of one relevant dimension enabled *Ss* to correctly identify half of the instances as negative, because half of the trials contained the negative level of either one of the relevant dimensions. The consistently correct selection of the positive instances, however, required *Ss* to take into account both of the relevant dimensions on a given trial. For these reasons, the number of correctly identified positive instances may be a more sensitive measure of the degree of complete solutions attained by *Ss* than is a total error or total correct measure. However, the number of correctly identified positive instances, will, on the average, increase as *S* increases the number of positive lever responses and because of this the increase in correct positive lever responses need have no direct relationship to solution of the problem. Therefore, the number of correct positive lever responses was corrected for the guessing behavior of *Ss*.

In the following analysis, *Ss*' scores, *X* (correct positives), were computed with the rule:

$$X = C^{+} - \frac{C^{-}}{3} \qquad [1]$$

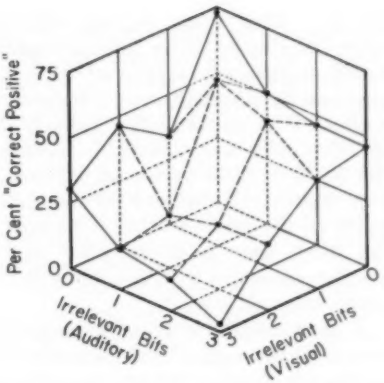

FIG. 3. Relationship between correct positives and amounts of irrelevant auditory and visual information. (Each plotted point represents the mean of six *Ss*. The origin of the vertical axis is at the chance level.)

where C^{+} is the number of correct positive lever responses (correctly identified positive instances) and C^{-} is the number of incorrect positive lever responses. The last term of Equation 1, a correction for guessing, was based on the assumption that the average distribution of "guesses" on the positive lever response will be 25% correct and 75% errors. The score, *X*, should then be a good estimate of the number of correct positives above chance performance.

The effects of the amount of visual and auditory irrelevant information on the degree to which *Ss* exceeded random behavior in the correct assignment of positive instances is shown in Fig. 3. As the amount of irrelevant visual information was increased from 0 through 3 bits, the mean percentage of correct positives decreased as follows: 53.9, 49.0, 29.0, and 15.3. The corresponding mean values for the auditory dimension were 45.8, 37.5, 34.9, and 29.1. Bartlett's test for homogeniety of variance among the cells gave a chi square of only 28.5 for

47 *df*. This value is well within accepted chance limits. An extended trend analysis of variance of the correct positives indicated that variation in both the auditory ($F = 5.58$, 1/48 *df*, $P < .05$) and visual ($F = 36.98$, 48 *df*, $P < .001$) irrelevant information significantly influenced the difficulty of the task in a linear manner. All other sources of variation gave results consistent with the analysis of error scores (Table 2).

The significant interaction between halves of the task and the amount of irrelevant visual information is shown in the right half of Fig. 2. It can be seen that the increase in correct positives from the first to the second half of the task became progressively smaller as the number of irrelevant visual bits was increased.

Solvers.—On the basis of the written solutions obtained from *S*s at the end of the session, it was quite easy to determine which *S*s had attained adequate verbal solutions to the problems. In addition, the verbal statements clearly indicated partial solutions, mostly the association of the negative level of one relevant dimension with negative instances. Only 36 of the 96 *S*s verbalized the solutions to the problems. The distribution of these solutions was in line with the previous analyses, with the number of solutions decreasing as the number of irrelevant dimensions was increased in either the auditory or visual signals.

Of the 60 *S*s who did not give correct verbal solutions to their problems, 22 correctly identified the relevant visual dimension (at least in its relation to negative instances) while only 4 *S*s did so for the relevant auditory dimension.

Discussion

The obtained linear function relating amount of irrelevant visual information and performance degradation is in line with most of the previous research. An exception is the Archer, Bourne, and Brown (1955) study in which task difficulty increased at a positively accelerated rate as irrelevant information was increased. However, they used stimuli which moved across a screen in a set period of time, whereas all of the other studies with this type of task involved stimuli which were present at least until *S* had responded. It is suggested that the action of a memory factor may account for the relative extreme difficulty in the high complexity conditions of Archer et al. (1955). Under the variety of conditions allowing unlimited observation of stimuli, the difficulty of concept identification is proportional to the amount (in bits) of irrelevant visual information in the stimuli, at least up to 5 bits.

The significant interactions of visual information and halves of the task indicate that the rate of learning decreases as the amount of irrelevant visual information increases. This is consistent with a significant interaction between trials and irrelevant visual information found by Brown and Archer (1956).

Perhaps the most noteworthy result in the present study was the lack of a significant effect (as measured by error scores) of variations in the amount of irrelevant auditory information. The predominance of the Type A errors over the Type V errors indicates that *S*s did not attend to the auditory cues as carefully as to the visual cues. It will also be recalled that this difference did not change significantly between the two halves of the task. It is therefore suggested that *S*s entered the situation with a bias toward using visual information in preference to auditory information and maintained this preference throughout training. This explanation is supported by the verbal reports of a few *S*s who volunteered the information that they had ignored the auditory signal during the session and by the small number of verbal identifications of the relevant auditory dimension.

The measure of percentage of correct positives was intended to provide an indication of the effect of the auditory information with a good deal of the proposed visual bias removed. The significant linear relationship between correct positives and amount of irrelevant auditory information is certainly the type of result expected on the basis of studies with visual stimuli.

The data offer no evidence that there is an interaction between irrelevant visual and auditory information.

Mathematical theories.—Application of statistical learning theories to the concept identification situation have been attempted by Bourne and Pendleton (1958) and by Walker (1958). The formulations of these authors are based on models proposed by Restle (1955, 1957).

The most extensive treatment of the concept situation to date is given by Bourne and Restle (1959). This approach incorporates the earlier applications and makes extensions to more complex situations. Bourne and Restle assume that the relevant cues are perfectly correlated with the reinforcements, i.e., have validity of 1, and they incorporate the cue adaptation model which Restle (1957) has developed.

The present study differs from prior work in that although *S* has only two choices of response he still must consider two stimulus dimensions for solution to the problem. In other words, the average validity of any one dimension is never 1 in the present situation. Restle (1957) has proposed, from empirical results, that the validity of a cue, *V*, be;

$$V = 4\pi^2 - 4\pi + 1, \qquad [2]$$

where π is the proportion of times a given cue is reinforced. Clearly, in the present study, cues from the positive level of a relevant dimension are reinforced on only half of the trials and thus have validity of 0. According to the model, such cues will adapt out of the effective sample. Also the negative level of either relevant dimension can serve only for the identification of negative instances. On this basis, a solution to the problem is not possible! On what basis then are the solvers solving the problem? A possible answer is consideration of cues arising from the combination of the two positive levels from the relevant dimensions (verbal hypotheses?). Such cues have validity of 1 since they are consistently reinforced. The quantitative development of this modification is difficult in view of the additional problems indicated below.

The lack of symmetry in the effect of irrelevant visual information versus irrelevant auditory information does not allow a simple application of the models mentioned above. In these theories, the proportions of effective cues originating from given stimulus dimensions have been taken to be equal (with one exception, color of the stimuli). However, the unequal effectiveness of auditory and visual signals forces some modifications in this assumption.

One additional problem has been created by the present procedure. The so-called irrelevant dimensions in the present study have some validity (assuming Equation 2), for the probability of reinforcement of a given cue from an irrelevant dimension is .25 and not .50 (which would be 0 validity). This problem raises the question of the appropriateness of the term "irrelevant" as used in this paper. Such dimensions are clearly capable of influencing behavior in the manner of the probabilistic matching solutions (Estes, 1950; Restle, 1957) but are not relevant to the *complete* solution possible in the concept task. The implication intended here is that to very sophisticated *S*s cues which allow a fairly high proportion of correct responses may be considered irrelevant when a perfect solution is possible, while to naive *S*s the same cues may be considered as relevant, perhaps as the best possible ones.

Summary

Performance in a concept identification task was studied as a function of the amounts of simultaneously presented auditory and visual irrelevant information.

Groups of 2 *S*s were randomly assigned

to each of 48 conditions. The design involved four levels (0, 1, 2, and 3 bits) of irrelevant visual information, four levels (0, 1, 2, and 3 bits) of irrelevant auditory information, and three problems (1 auditory and 1 visual bit of relevant information in each problem).

The *Ss* were instructed to assign each stimulus (consisting of both an auditory and a visual portion) to one of two categories by depressing one of two lever-action switches.

The major results of the study were: (*a*) As the amount of irrelevant visual information was increased, performance decreased in a linear manner. (*b*) The variation in the amount of irrelevant auditory information did not significantly affect the number of errors. (*c*) The number of correctly identified positive instances of the concept (a measure of the degree of complete solution) significantly decreased as a linear function of the amount of irrelevant visual information and as a linear function of the amount of irrelevant auditory information. (*d*) The types of errors made by *Ss* gave strong evidence that *Ss* responded primarily on the basis of the visual stimuli and tended to ignore the auditory stimuli.

References

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

Bourne, L. E., Jr. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

Bourne, L. E., Jr., & Pendelton, R. B. Concept identification as a function of completeness and probability of information feedback. *J. exp. Psychol.*, 1958, **56**, 413–420.

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

Brown, F. G., & Archer, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

Estes, W. K. Towards a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

Restle, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.

Walker, C. M. Concept identification as a function of amounts of relevant and irrelevant information. Unpublished doctoral dissertation, University of Utah, 1958.

(Received July 12, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 291–294

EYELID CONDITIONING PERFORMANCE AS A FUNCTION OF EMOTION-PRODUCING INSTRUCTIONS[1]

KENNETH W. SPENCE AND HENRY GOLDSTEIN

State University of Iowa

A series of recent studies from the Iowa laboratory has shown that performance level in eyelid conditioning of humans is related to (*a*) UCS intensity (Spence, 1953, 1958; Spence & Taylor, 1951), (*b*) extreme scores on a questionnaire designed to select *S*s differing in the degree to which they manifest emotional symptoms (Spence & Farber, 1953; Spence & Taylor, 1951; Taylor, 1951), (*c*) physiological indices of emotionality of *S*, such as GSR, heart rate, and muscle action potential (Runquist & Ross, 1958, 1959; Runquist & Spence, 1959), and (*d*) shock and/or threat of shock given *S* prior to or during the course of conditioning (Spence, 1958; Spence, Farber, & Taylor, 1954). These findings have been interpreted as reflecting different levels of drive (*D*) which in turn has been conceptualized in terms of different levels of a hypothetical emotional response (r_e) in *S* (Spence, 1958).

As a further test of this theory that the level of the emotional response of *S* determines *D* and hence response strength ($R = fE = H \times D$), an experiment was conducted in which *S*s, after receiving 20 conditioning trials with a weak UCS (air puff to cornea) were informed that the UCS was going to be changed to a very intense value that would be quite unpleasant. On the assumption that such a threat would markedly increase the level of emotional response it was predicted that the conditioning performance of this experimental group would be higher than that of a control group not so threatened. A second experimental group was run which not only was threatened with a stronger puff after Trial 20, but actually received the stronger puff from this point on. Comparison of these two experimental groups provided a measure of the relative effects on performance of the threat alone and the threat plus increased intensity of the puff.

METHOD

Apparatus and method of recording.—The equipment for recording eyelid closure and presenting the UCS was identical with that used in previous studies (cf. Spence & Taylor, 1951). The CS was an increase in brightness of a 6-cm. circular disk from .004 to 2.95 mL. The weak UCS was a 0.33-psi air puff and the strong UCS a puff of 3.0 psi. The UCS was delivered to the right eye and was limited in duration to 50 msec. The CS-UCS interval was 500 msec. with the duration of the CS being 550 msec. A CR was recorded whenever the record showed a deflection of 1 mm. or more in the interval 200–500 msec. following the onset of the CS. Responses with a latency of less than 200 msec., which were infrequent, were classified as original responses, and were not included in the data.

Procedure.—Each *S* was instructed to blink once to a ready signal (the word "ready") and then look at the disk, remaining as relaxed as possible and making no attempt to control the movements of his eyelid. Following the reading of the instructions each *S* received three presentations of the CS alone followed by a single presentation of the UCS. All *S*s were then given 40 conditioning trials. Intertrial intervals of 15, 20, or 25 sec., averaging 20 sec. and arranged by a fixed

[1] This research was carried out as a part of a project concerned with the influence of motivation on performance in learning under Contract N9 ONR-93802, Project NR-154-107 between the State University of Iowa and the Office of Naval Research. Acknowledgement is made to David Lucky who ran some of the *S*s.

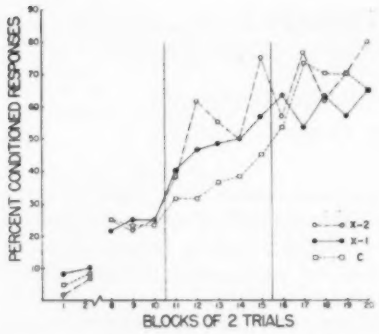

FIG. 1. Mean percentage of CRs in blocks of two trials.

schedule, were used. The interval between the ready signal and CS onset was 2, 3, or 4 sec., randomly varied.

The three groups were differentiated in terms of both the number of the trial after which a threat of the strong UCS was given and the number of the trial on which the strong UCS was actually introduced. All groups received the weak (0.33-psi) puff for the first 20 trials. The control *Ss* (Group C) were given 10 further conditioning trials with the weak UCS and then immediately after Trial 30 were threatened with and given the strong UCS for the next 10 trials (31–40). The first experimental group (X-1) received the threat of the strong UCS immediately after Trial 20, but the weak UCS was continued until Trial 30 at which time the strong UCS was introduced and used to the end of conditioning (Trial 40). The second experimental group (X-2) received the threat after Trial 20 with the strong UCS also being introduced immediately (i.e., on Trial 21) and continued until the end of training (Trial 40).

While the critical data in which we were interested was performance on Trials 21–30, the extra 10 trials were run for two reasons. First, so that all *Ss* would be shifted to a strong puff and would so report if they discussed the experiment with others. The second reason was that we wished to eliminate so-called "inhibitors," *Ss* who showed little or no conditioning even with the strong UCS and Trials 31–40 offered a period in which the drive condition (intensity of UCS) was the same for all *Ss*.

The verbal threat consisted, essentially, of telling each *S* that following one of the next 10 presentations of the light (CS), and following every presentation thereafter, he would receive a very strong and unpleasant puff of air (UCS). It was stressed that the first presentation of the stronger puff would be determined by chance (by drawing a number out of a box) and that *S* would have no way of knowing ahead of time on which of the next 10 trials he would actually begin receiving the strong puff. At the conclusion of the session *S* was questioned as to his knowledge of the purpose of the experiment and warned not to discuss the experiment with other members of the class.

Subjects.—The *Ss* were 51 men and 39 women students from a course in elementary psychology. At the beginning of the experiment each *S* was assigned to one of three groups in order of appearance at the laboratory. Towards the end of the experiment, *Ss* were assigned to the different groups so as to match as closely as possible their average performances on Trials 1–20. Each group then consisted of 30 *Ss*, 13 females and 17 males. A number of additional *Ss* were discarded for various reasons, 11 because they met the previously established criterion defining a voluntary responder (Spence & Ross, 1959). Nineteen were eliminated as "inhibitors" because they failed to give at least two CRs in the final block of 10 trials, during which all *Ss* received the strong UCS. One *S* was eliminated because he gave CRs during the preconditioning trials.

RESULTS AND DISCUSSION

Figure 1 shows the performance curves in terms of the percentage frequency of CRs of the three groups of *Ss* for each block of two trials over the course of conditioning. As may be seen, the groups were very closely matched on the first 10 blocks of trials. The differences in performance levels on Blocks 11–15 reflect

TABLE 1

MEAN INCREASE IN PERCENTAGE OF CRS FROM BLOCK OF 10 TRIALS PRECEDING THE INSTRUCTIONS TO BLOCK OF 10 TRIALS FOLLOWING

Group	Mean Percentage Increase	*SD*
C	12.7	3.94
X-1	25.0	4.20
X-2	32.0	3.52

the effects of the different experimental procedures for manipulating level of emotionality. Thus it will be seen that the curve of Group X-2, which was threatened with and actually administered the strong (3-psi) puff on Trials 21–30 jumped up the most. The performance of the group (X-1) only threatened with a strong puff also showed marked acceleration and, as may be seen, its performance was substantially above that of the control group (C) on this block of trials.

Table 1 presents both the mean percentage of CRs from Trials 11–20 to Trials 21–30. Evaluation of the differences between these increases by a simple analysis of variance provided an F of 7.41 ($P < .005$). A comparison of the difference between Group C and Group X-1 yielded a t significant at the .025 level, while that between Group C and Group X-2 was significant at the .001 level. The difference between the two experimental groups was not significant ($P = .22$).

These findings clearly support the prediction based on our emotional response theory of drive (D) that the level of conditioning performance of the experimental group (X-1) threatened with a strong, unpleasant puff would be significantly higher than the control group (C). In connection with this finding it is interesting to recall the results of Taylor (1951). In a study primarily concerned with the relation between conditioning performance and the level of emotionality of *Ss* as determined by a questionnaire-type scale, Taylor also attempted to manipulate experimentally the emotionality level of her *Ss* after 20 conditioning trials. Half of each type of her *Ss* was instructed that the intensity of the UCS would be increased on subsequent trials and half told that the intensity would be decreased. Actually the intensity of the UCS remained unchanged.

TABLE 2

COMPARISON OF RESULTS OF PRESENT STUDY WITH THOSE OF TAYLOR (1951)

Study	Group	Mean Percentage Increase of CRs	
		Trials 21–25 vs. 11–20	Trials 26–30 vs. 11–20
Present	C	8.7	16.7
Present	X-1	21.3	28.7
Taylor	Plus	17.0	11.4

Taylor found that the mean increase in percentage of CRs for Trials 21–30 over Trials 11–20 was 14.2 for her two groups (high and low emotionality *Ss* combined) that were threatened with an increase in puff intensity. This is to be compared with 25% in the case of our threatened group. Suspecting that the effects of her threat instructions had only a temporary effect, Taylor further analyzed her data and found corroborative evidence in that the increase over Trials 11–20 for this threatened group was, as shown in Table 2, 17.0% for Trials 21–25, whereas in Trials 26–30 there was a drop in performance to a level only 11.4% above that of Trials 11–20.

Anticipating the development of this scepticism of a change in puff intensity on the part of *S*, we attempted to prevent its occurrence by framing our instructions so that *S* would continue to expect the change even when it was observed not to occur over the 10 trials. That our procedure was successful in this respect is revealed by the fact that our threatened group (X-1) continued to show an increase in the second block of five trials. Thus, as revealed in Table 2, Group X-1 showed a larger increase over Trials 11–20 on Trials 26–30 (28.7%) than on Trials 21–25 (21.3%). As may be seen by comparing these latter values with the comparable ones for the control group (C), the amount by which the performance of the threatened group was elevated over that of the control group is approximately the same in both blocks of trials (12.6 vs. 12.0).

Summary

This experiment was conducted to ascertain the effects of introducing emotion-elevating instructions during the course of an eyelid conditioning experiment. One group of *Ss* that had received 20 conditioning trials with a weak UCS (.33-psi puff) was informed that the UCS would be increased to a very unpleasant intensity at some time during the following 10 trials. On the assumption that this threat would materially increase the level of emotional response (r_e) and hence the level of D of S, it was predicted that the performance of this group over this block of trials would be significantly higher than that of a control group not given such a threat. This prediction was confirmed. A third group was run in which the UCS was actually changed to a rather strong puff (3.0 psi) after 20 trials with the weak puff. The performance level of this group on the test trials (21–30) was slightly, but not significantly, above that of the threatened group. Thus the drive-increasing effect of the instructions was slightly less than that produced by increasing the UCS from a puff of .33 psi to one of 3.0 psi.

References

Runquist, W. N., & Ross, L. E. A preliminary investigation of the relationship between GSR, heart rate change, and eyelid conditioning. *Proc. Iowa Acad. Sci.*, 1958, **65**, 398–404.

Runquist, W. N., & Ross, L. E. The relation between physiological measures of emotionality and performance in eyelid conditioning. *J. exp. Psychol.*, 1959, **57**, 329–332.

Runquist, W. N., & Spence, K. W. Performance in eyelid conditioning related to changes in muscular tension and physiological measures of emotionality. *J. exp. Psychol.*, 1959, **58**, 417–422.

Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

Spence, K. W. A theory of emotionally based drive and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

Spence, K. W., Farber, I. E., & Taylor, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.

Spence, K. W., & Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

Spence, K. W., & Taylor, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

Taylor, J. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

(Received July 16, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 295–301

PROACTIVE INHIBITION OF CONNECTED DISCOURSE[1]

NORMAN J. SLAMECKA

University of Vermont

The phenomenon of proactive inhibition (PI) has never been reported for the rote retention of prose materials (Slamecka & Ceraso, 1960). The main purpose of this paper is to determine whether passages of connected discourse are in fact subject to such PI, and also to examine the effects of two variables upon the magnitude of such interference. One of these variables is the amount of prior learning (PL), expressed in terms of the number of interfering passages learned. With paired associates, Underwood (1945) found that PI increased directly with increases in the number of prior lists learned. The other variable under investigation is the interval between the learning and test for retention of Passage 2. According to logical expectations, PI should increase with time (up to some undetermined point), and the available evidence suggests that this is probably the case, at least up to 48 hr. (Slamecka & Ceraso, 1960). Since our previous studies have indicated the generalizability of findings based upon nonsense materials to prose passages (Slamecka 1960a, 1960b), we made the following hypotheses about the present experiment: (*a*) the rote retention of passages of connected discourse is subject to significant PI, (*b*) PI varies directly with the number of previous passages learned, and (*c*) PI increases with increasing retention intervals.

[1] This research was sponsored by National Science Foundation Grant G-6192. The author is indebted to John Corson and to William Adams, who assisted in the collection and analysis of the data.

In addition, this study is also concerned with another aspect of knowledge about retention, strikingly pointed out by Underwood (1957). By summarizing retention data from a number of studies, he showed that the percentage recalled after 24 hr. decreased directly as a function of *S*'s stage of practice, or the number of lists previously mastered. Thus, PI effects (through the use of highly practiced *S*s) have served to depress the amount recalled below that actually to be expected with naive *S*s, and the commonly published retention curves found in textbooks are confounded with such effects. Underwood (1957) concludes that "we must start all over and derive a retention curve over time when the subjects have learned no previous materials in the laboratory" (p. 56). To our knowledge there is in the literature no such curve based upon connected discourse. Underwood concludes that the loss in retention of verbal materials is only 25% over 24 hr. with naive *S*s, and that even this value is inflated by the failure to take into account the postcriterional drop in performance at the conclusion of learning. The present study shall provide data for four retention curves covering a 24-hr. period, one of which is a curve produced by naive *S*s, allowing an empirical comparison of Underwood's findings to connected discourse.

METHOD

Subjects and materials.—Thirty-six students in general psychology at the University of Vermont served as *S*s as part of the requirements of the course. Every *S* participated

in four individual sessions, during each of which he memorized a prose passage and was tested for subsequent retention after a predetermined interval. The materials were four fairly difficult sentences, each exactly 20 words in length, all rather similar in subject matter. The passages used were:

1. "Any generalized statement of relationship between variables requires specification of some reference class or universe from which the material comes."

2. "According to recent considerations it appears preferable to select preliminary assumptions from the realm of objectively verifiable and observable phenomena."

3. "Traditional rationalism cannot negate the assertion that its theory concerning primary evidence rests upon inadequate analysis of fundamental scientific methodology."

4. "Forthright admission of the contingent character of our most soundly established beliefs distinguishes contemporary empiricists from dogmatic and positive ancients."

The words of the passages were presented on a memory drum at a 3-sec. presentation rate and a 6-sec. intertrial interval. Learning was by serial anticipation.

Design and procedure.—A mixed factorial design (Lindquist, 1953, Type II) with four levels of each variable was used. One variable was the number of prior passages learned (0, 1, 2, and 3), and the other was the retention interval (15, 30, and 60 min., and 24 hr.). Of the 16 combinations of these levels, each *S* had four, in such a manner that he had all levels of each variable but no repetitions of any level. The 36 *S*s provided 9 replications of the 16-cell matrix. Conditions were arranged so that passages were counterbalanced with respect to the two main variables. Sessions, of course, were counterbalanced only with respect to retention intervals and passages but not with amount of PL, since the session sequence determined the PL level, i.e., Sessions 1, 2, 3, and 4 were always associated, respectively, with 0, 1, 2, and 3 previous passages.

Each experimental session began with one warm-up trial of number guessing on the memory drum. Next, instructions for serial anticipation were given and then the designated passage was presented. The criterion for original learning (OL) was one errorless trial plus one additional trial.[2] The designated retention interval was passed outside of the laboratory. For the 15-, 30-, and 60-min. intervals *E* engaged *S* in casual conversation, and for the 24-hr. interval *S* was dismissed until the following day. Upon return to the laboratory, *S* was given one trial of number guessing for warm-up, followed by relearning of the last passage. The criterion for relearning was one errorless trial, but all *S*s were given a minimum of five trials, regardless of whether they may have relearned in fewer than that. As soon as relearning of one passage was complete, *S* was given the next passage to learn, and the above procedure was repeated until each *S* had successively acquired and relearned all four passages.

Results and Discussion

The acquisition data were evaluated by way of three criteria. Mean total correct anticipations per *S* are presented in the upper half of Table 1, and the results of the statistical analysis are found in the left half of Table 2. It is seen that amount of PL had a significant facilitating effect in that the better practiced *S*s needed fewer reinforcements to learn the passages.

Acquisition performance was also measured by total trials during learning, and by performance level on the last trial. Mean total trials as a function of increasing PL levels were, 14.2, 11.8, 10.5, and 10.1. These are significant beyond the .001 level ($F = 22.52$, $df = 3/96$). Mean numbers correct on the last OL trial (the trial following the errorless criterion trial) with increasing amount of PL were 18.61, 19.11, 19.39, and 19.42. The expected postcriterional drop did occur, and although the differences were slight the means were significant beyond the .05 level ($F = 3.44$, $df = 3/96$). All three analyses are consistent in showing a significant

[2] This would provide information on the postcriterional drop for prose materials. Hilgard (1951) says: "Any criterion that depends upon the first reaching of some high level of performance capitalizes on chance fluctuations in the direction of good scores. The mean of scores just after the criterion will tend to fall below the criterion trial" (p. 554).

TABLE 1

MEAN TOTAL CORRECT ANTICIPATIONS DURING ACQUISITION AND MEAN RECALL SCORES

Prior Learning (Lists)	Retention Intervals							
	15 min.		30 min.		60 min.		24 hr.	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
	Acquisition							
0	214.3	93.7	190.6	50.4	179.1	39.7	160.8	55.8
1	125.7	28.2	134.3	45.2	153.7	56.2	189.8	34.8
2	147.3	39.6	121.7	40.5	170.3	76.8	138.2	50.4
3	124.0	25.9	153.6	30.9	132.6	35.0	131.9	44.7
	Recall							
0	18.1	1.2	17.9	1.9	17.8	1.9	15.2	3.2
1	18.1	2.4	16.8	1.5	15.4	2.7	13.3	3.1
2	15.2	3.3	16.1	3.2	13.7	3.6	11.7	3.5
3	13.7	3.6	13.3	2.9	13.2	3.4	9.3	4.9

practice effect for acquisition of prose materials, and are in good accord with similar facilitating effects commonly reported for nonsense items. None of the groups differed on original learning as a function of the subsequent retention intervals.

The main analysis of retention was through the recall (first relearning trial) scores. Recall data are presented in the lower half of Table 1, and the statistical analysis summary is in the right half of Table 2. The effects of amount of PL and retention intervals were significant. Appropriate *t* tests indicated that all adjacent means differed from each other at or beyond the .025 level for the PL variable, and with retention interval only the 24-hr. group differed significantly from all others ($P < .01$). Because of the possibility of heterogeneity of variance in these data, supplementary analyses were performed using the distribution-free Friedman analysis of variance of

TABLE 2

ANALYSES OF VARIANCE FOR TOTAL CORRECT ANTICIPATIONS DURING ACQUISITION, AND FOR MEAN RECALL SCORES

Source	Acquisition			Recall		
	df	*MS*	*F*	*df*	*MS*	*F*
Between *Ss*	35	6893.8	.99	35	12.1	.94
PL × Int. (b)	3	6136.1	.88	3	3.9	.31
Error (b)	32	6964.8		32	12.8	
Within *Ss*	108	2322.9	1.35	108	18.1	1.55*
PL	3	25572.7	14.90**	3	160.8	13.75**
Interval	3	489.1	.28	3	113.8	9.74**
PL × Int. (w)	6	1313.6	.76	6	1.8	.16
Error (w)	96	1716.8		96	11.7	

* $P < .05$.
** $P < .01$.

TABLE 3

RELATIONSHIP BETWEEN ACQUISITION RATES AND RECALL

Prior Learning (Lists)	Total Correct Anticipations per *S* in Acquisition							
	50–99		100–149		150–199		200-more	
	N	Mean	*N*	Mean	*N*	Mean	*N*	Mean
0	0	—	12	17.3	12	17.2	12	17.2
1	7	16.9	11	16.6	11	15.7	7	14.1
2	10	13.0	11	16.1	10	14.3	5	12.0
3	9	11.8	13	12.3	12	13.3	2	10.0

ranks. A χ_r^2 of 34.3 for PL ($P < .001$, $df = 3$), and of 20.4 for retention interval ($P < .001$, $df = 3$), confirmed the effectiveness of both variables.

At this point in the analysis the problem of the role of practice effects upon retention had to be examined. It could be argued that the differences at recall were not true PI effects, but rather were the result of differential acquisition rates. That is, *S*s who needed more total trials or correct anticipations during acquisition of a passage should recall it better than those who took fewer trials, simply because of their more intensive training upon it. Two approaches were used to investigate this possibility. First, within each PL level the correlation between total correct anticipations and recall was calculated. If the degree of specific training on a passage did indeed determine its retention, there would be a significant positive correlation, and the recall scores would have to be adjusted accordingly. The obtained *r*'s with increasing PL levels were respectively, .09, −.16, −.19, and .06 (all at $df = 35$). None of these approached significance, thereby indicating no need for adjustments and no evidence for a within-PL groups effect.

The second method, originally used by Underwood (1945), involved an across-PL groups comparison of recall, with each PL group subdivided into categories of total correct anticipations (or total reinforcements). The relevant data are found in Table 3. If it were the case that recall level was determined by acquisition rate, then the recall means would rise from left to right across the table, and also would remain constant vertically, across PL levels. Inspection of the table shows that this is not so, but rather that within each total correct anticipation category the recall means decrease steadily with increasing PL levels as would be expected of true PI effects.

These data are also consistent with Underwood's (1954) conclusion that fast and slow learners do not differ in retention. At first glance it may appear, paradoxically, that amount of reinforcement during acquisition has no effect upon retention, but this is only the result of our sorting out *S*s on the basis of learning rate. Studies specifically designed to demonstrate the effects of various degrees of learning upon retention (Briggs, 1957; Slamecka, 1960a) usually employ a fixed trials acquisition procedure wherein each treatment group is presumably heterogeneous with respect to fast and slow learners.

From the two preceding analyses we conclude that the recall data prove the presence of genuine PI, which confirms the main hypothesis of this study, and that the degree of PI increases with increasing numbers of previously learned passages, thus confirming the second hypothesis. Expressed in terms of the formula, (Rest-Work/Rest) × 100, relative PI as a function of increasing amounts of PL was found to be 8.0%, 18.1%, and 28.6%. Relative PI as a function of increasing retention intervals was 13.5%, 13.9%, 20.6%, and 24.8%, thus confirming the third hypothesis that PI increases with time, at least up to 24 hr.

Analysis of total trials to relearn showed the usual transitoriness of PI in that there was no significant effect of amount of PL ($F = .85$, $df = 3/96$),

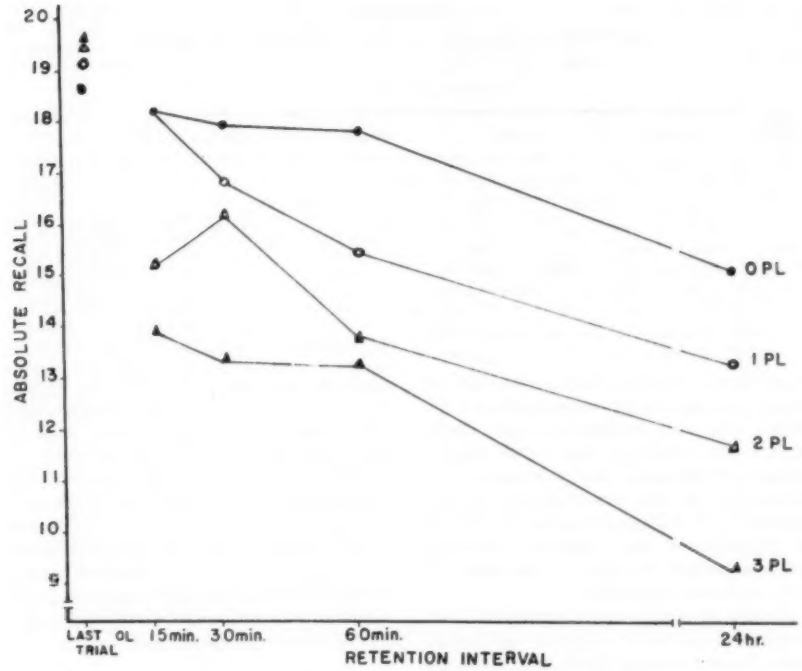

FIG. 1. Absolute amount recalled as a function of amount of prior learning (PL) and retention interval.

but retention interval was significant beyond the .01 level ($F = 5.87$, $df = 3/96$).

Retention curves in terms of absolute amount recalled, together with the terminal OL trial performance of each group are found in Fig. 1. The uppermost line is the curve for naive *S*s. In spite of the fact that recall on the postcriterional trial was better the greater the amount of PL, and that all of the long-duration recalls are inflated somewhat by the learning on that last trial, the orderly decremental effects of increasing PI are clearly evident at each retention interval. These results strongly suggest that the classical Ebbinghaus type of curve, with its steep initial drop and considerable overall decrement (most closely approximated by our 3-PL group) is a function of cumulative proactive interference attributable to practiced *S*s.

Underwood (1957) estimated that naive *S*s actually forgot only 15% over 24 hr., with the assumption of a postcriterional drop of about 10%. Our data show that the mean postcriterional drop for naive *S*s was 1.39, or 7%, which is in fair accord with Underwood's figure. Then, using the values of 18.61 (mean performance of naive *S*s on final learning trial) and 15.22 (mean recall of naive *S*s after 24 hr.) we obtain a retention loss of 18.2%, which is a slight underestimate because of the unmeasured learning during the postcriterial trial. Considering the variability resulting

TABLE 4

PERCENTAGES OF OMISSION AND INTRUSION ERRORS AT RECALL

Prior (Lists)	Retention Interval								Mean	
	15 min.		30 min.		60 min.		24 hr.			
	Om.	Int.	Om.	Int.	Om.	Int.	Om.	Int.	Om.	Int.
0	5.6	3.9	6.1	4.4	8.3	2.8	17.8	6.1	9.4	4.3
1	7.2	2.2	13.3	2.8	16.7	6.1	22.8	10.6	15.0	5.4
2	19.4	4.4	14.4	5.0	21.7	10.0	35.6	6.1	22.8	6.4
3	26.1	5.6	26.7	6.7	27.8	6.1	46.7	6.7	31.8	6.3
Mean	14.6	4.0	15.1	4.7	18.6	6.2	30.7	7.4		

from the many sources utilized by Underwood, the degree of correspondence between the two findings is sufficiently high. We conclude that even though connected discourse may possess unique properties, its rote retention and its susceptibility to PI as a function of amount of interference and retention interval are comparable to those of unconnected and nonsense materials.

Errors at recall were categorized into omissions (failure to respond) and intrusions (source unspecified), and the percentage occurrence of each is given in Table 4. Examination of the table shows that intrusions played a very minor role throughout, and that they increased only slightly with increases in interference levels and retention intervals. As in previous studies (Slamecka, 1960a, 1960b), omission errors accounted for the bulk of the changes at recall, for they increased markedly with increases in PL and with lengthening retention intervals. This again points to the hypothesis that with serial lists of connected discourse material, the main effect of interfering lists is to increase the reaction latencies of the correct resonses, and with the use of limited response times (3 sec. in this study), the obtained omission pattern would follow. Further support for the contention that interference affected mainly the latencies but not the actual retention of the items in given by calculating the percentage of all overt responses given at recall, that were correct. These accuracy scores, in order of increasing PL levels, are: 95.2%, 93.6%, 91.7%, and 90.9%, and suggest that approximately the same percentage of responses attempted was correct. It would seem that with this material, the importance of intrusions as indicators of degree of interference is overshadowed by the omission pattern, and that future work might fruitfully be directed toward obtaining actual response latencies under conditions of unlimited recall time, as a test of the above hypothesis.

SUMMARY

This study tested the susceptibility of the rote retention of prose passages to proactive inhibition (PI) of recall, as a function of amount of prior learning (PL) and length of retention interval. Thirty-six *S*s each learned and recalled four 20-word passages of connected discourse in succession, by the method of serial anticipation. There were four levels of amount of PL (0, 1, 2, and 3 passages), and retention intervals were 15, 30, and 60 min., and 24 hr. The following results were obtained: (*a*) Significant practice effects occurred during acquisition. (*b*) Significant PI of recall was obtained. (*c*) PI increased directly with the amount of PL. (*d*) PI increased directly with the retention interval. (*e*) After 24 hr. naive *S*s forgot about 18%, thus supporting Underwood's findings about this point. (*f*) Omission errors at recall were more predictive of the retention pattern than were intrusions.

It was concluded that whatever unique properties may be possessed by connected discourse, it also shows the same susceptibility to PI effects as does nonsense material. A hypothesis that interference of this kind

serves mainly to increase the latency but not the accuracy of correct responses, was discussed.

REFERENCES

BRIGGS, G. E. Retroactive inhibition as a function of degree of original and interpolated learning. *J. exp. Psychol.*, 1957, **53**, 60–67.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of practice level. *J. exp. Psychol.*, 1960, **59**, 104–108. (a)

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of similarity of topic. *J. exp. Psychol.*, 1960, **60**, 245–249. (b)

SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 273).

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

(Received July 18, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 302–309

RETENTION OF INFORMATION UNDER CONDITIONS APPROACHING A STEADY STATE[1]

ROGER N. SHEPARD AND MARTHA TEGHTSOONIAN
Bell Telephone Laboratories *Eastern Pennsylvania Psychiatric Institute*

In experiments on human memory, *S*s are typically presented with an isolated block of information and then tested for retention of that information. The digit span test is a familiar example. In experiments of this kind *S*s have recourse to various devices of covert regrouping, recoding, and rehearsal of the information. Furthermore, interference resulting from prior exposure to similar information is usually minimized. The results of such experiments may not, therefore, be generalizable to situations that confront *S*s with a continuing sequence of items and require them, throughout the sequence, to retain as much as possible of the most recently presented information. Situations of this type are nevertheless quite common (as, for example, in playing a hand of bridge) and in some cases they are of considerable practical importance (as in the continuous monitoring of a complex display). They also have a wide theoretical bearing; for in such varied activities as inductive inference, deductive reasoning, problem solving, and verbal communication *S*s' responses at any one time must be selectively modulated by stimuli variously strung out into the recent past. Cahill and Hovland (1960) have demonstrated the crucial role that this kind of memory plays during the acquisition of concepts.

Recent efforts to measure retention during continuous tasks have employed recall procedures (e.g., Mackworth, 1959; Pollack, Johnson, & Knaff, 1959). In the present experiment, however, a recognition procedure is selected instead. Unlike the usual recall procedure, the recognition procedure enables *E* to single out, for test, the one item presented at any specified earlier point in the sequence and so to vary at will the number of items intervening between the initial and test presentation of an item.[2] By testing after different numbers of intervening items, then, the course of forgetting of single items during the presentation of subsequent items can be inferred. A curve of forgetting can thus be obtained under conditions that depart from those usually imposed in that the interval between the original and test presentations of an item, far from being filled with an

[1] The present experiment was conducted at Harvard University under Contract AF33(038)-14343 with the Operational Applications Laboratory of the Air Force Cambridge Research Center. Reproduction for any purposes of the United States Government is therefore permitted.

This work was reported in April 1959 at the Annual Meeting of the Eastern Psychological Association. The authors are indebted to George A. Miller who furthered this work by encouragement and specific suggestions.

[2] Since the present report was originally prepared several other studies have appeared that use various ingenious methods to achieve a similar result with recall procedures (Anderson, 1960; Lloyd, Reid, & Feallock, 1960; Yntema & Mueser, 1960). The last two of these three studies used continuous tasks that, except for the utilization of a recall rather than a recognition test, were quite similar to the task used here. Indeed, Yntema and Mueser present curves of forgetting that seem to be the analogues, for the recall process, of the curve illustrated in the present Fig. 1.

entirely different activity, is filled simply with further presentations of exactly the same kind. In this way an extreme can be approached in which the possibility of rehearsal is minimized while the interference of preceding material is maximized.

In addition to making an empirical determination of the curve of forgetting under such conditions, there are two other aims: First, to show that a rational account can be given for these data in terms of a stochastic model originally developed to account for the shape of the gradient of stimulus generalization. And, second, to estimate the total amount of information (about all preceding stimuli of the sequence) that is carried along when steady state is reached; i.e., when the amount of information gained from each successive item just compensates for the amount forgotten from all preceding items.

Method

Subjects

The *S*s were 63 students at Harvard University. Of these, the 15 serving in the "variable delay condition" were graduate students in psychology, and the 48 serving in the "fixed delay condition" were summer session students who were paid for their participation.

Procedure

At the beginning of each experimental session *S* was given a large deck of 3 × 5-in. cards each bearing a single three-digit number. Following a brief description of the manner of construction of the deck, *S* was given a specially prepared record sheet and instructed to examine in turn the number on each card and to place a mark on the record sheet either in a column labeled "old" or in a column labeled "new" according as he did or did not remember having seen that number on an earlier card of the deck. He was allowed to proceed through the deck at his own rate but was not permitted to look back to any earlier card. No feedback as to the correctness of his responses was provided.

The three-digit numbers included in each deck were randomly selected from the total population from 000 through 999. However, with the exception of a few nonduplicated numbers toward the end of the deck, every number selected for a given deck occurred exactly twice in that deck. The second occurrences of the numbers were inserted in such a way that (after the first few cards) each successive card bore either an old or a new number with equal probabilities and according to a random sequence. The insertion of a few nonduplicated numbers toward the end was designed to prevent the probability of finding a new number from falling below .5 for the last cards.

The dependent variable was the probability of classification of the number on a given card as "old." It thus became the probability of a correct response, if the number on the given card had also appeared on an earlier card, and the probability of an incorrect response (or "false alarm"), if the number had not. There were two independent variables. The secondary independent variable, T, was simply the total number of other cards preceding a given card in the same deck. The primary independent variable, d ("delay"), applied strictly only when the number appearing on a given card had appeared on an earlier card. In that case it was defined as the number of other cards intervening between the given card and the earlier card bearing the same number. (Since the rate of presentation was not controlled, delay, as defined here, had only a rough relation to elapsed time. No attempt was made to separate experimentally the effect of the number of intervening presentations per se from the effect of pure temporal delay.) The number of intervening presentations was manipulated differently under the two following experimental conditions:

Variable delay condition.—Four decks were prepared for this condition. In each of these decks the delays varied from 0 to about 60 intervening cards for different three-digit numbers within the same deck. Insofar as possible a roughly rectangular distribution of delays was maintained throughout the decks. (Of course the delays were limited from above for old numbers occurring near the beginning of a deck.) Two of the decks, both containing 100 cards, were each given to 6 *S*s. The other two decks, both containing 200 cards, were each given to 10 *S*s. Although some of these *S*s served more than once, their exposure to different decks was always separated by 2 or more days. The results, moreover, failed to provide evidence of any interdeck transfer.

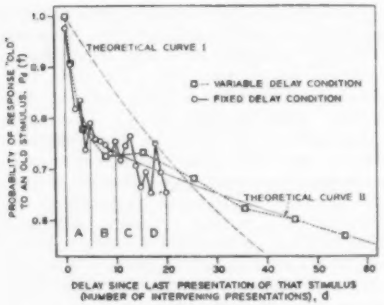

FIG. 1. Probability of classification of an old three-digit number as "old" as a function of the number of intervening three-digit numbers. (Each of the four decks of the variable delay condition was weighted, in calculating the average points, by the number of *S*s times the number of cards for that deck. Of the 21 points for the fixed delay condition, the 18 falling inside one of the four ranges of delays—A, B, C, or D—are each averaged from the two decks covering that range and the 3 points at the intersection of two adjacent ranges are each averaged from the four decks overlapping at that delay. The smooth curves in this and subsequent figures are based upon theoretical considerations presented in the theoretical analysis and discussion section.)

Fixed delay condition.—Eight decks were prepared for this condition. In each of these decks the delays never varied more than 2½ cards from the mean delay for the deck. Different decks realized different mean delays with approximately rectangular distributions for the following ranges: from 0 to 5 for Deck A_1, from 5 to 10 for B_1, from 10 to 15 for C_1, and from 15 to 20 for D_1. The number of cards in the decks varied from 200 to 217, but the number of different three-digit numbers that occurred twice in a deck was held constant at 100. Each of 24 *S*s went through all four of these decks according to a different one of the 24 possible orders. This complete counterbalancing provided for a separation of the effects of interdeck transfer and mean delay. The administration of successive decks to the same *S* was always separated by at least 1 day and in most cases by 2 or more days.

In a replication, 24 new *S*s went through Decks A_2, B_2, C_2, and D_2 according to the same counterbalanced design. These four new decks were prepared in the same way as the first four decks. The independent construction of new decks for the replication, however, insured a more adequate sampling of possible sequences of new and old numbers. In the fixed delay condition the probabilities of new or old numbers were necessarily not completely independent for neighboring cards. The restriction of the delays to a range of five cards for any one deck entailed small sequential dependencies. However, since the variable delay condition—for which there were no sequential constraints—yielded the same results, it seems unlikely that *S*s made any appreciable use of these constraints.

RESULTS

Effect of experience with previous decks.—The performance of *S*s on a given deck did not appear to depend upon the number of other decks to which they had already been exposed. In the fixed delay condition, the percentage of the total classifications (as "old" or "new") that were correctly made by *S*s ($N = 48$) in the four sessions, in the order of their occurrence, were 76, 76, 76, and 77% correct. Since practice evidently had little effect, the data from all four sessions were pooled for subsequent analyses.

Effect of delay, d.—Figure 1 summarizes the effect of delay upon the accuracy of classification of an old three-digit number as "old." The two methods of varying delay yielded essentially indistinguishable results. The probability that a number was classified as "old" dropped rapidly from almost 1.00 (when that number was just seen on the immediately preceding card) to about .75 (when that number had not been seen for 6 or 7 cards). From there on this probability dropped more slowly to about .57 (when the number had not been seen for as many as 50 or 60 cards). However, even for the longest delays imposed, this probability was well above the probability that a *new* number was classified as "old" (the

"false alarm" probability), which averaged only about 0.22.

Effect of number of cards already examined in the same deck, T.—Figure 1 might at first be thought to represent the course of forgetting of a single three-digit number under steady state conditions. However, comparisons between the performances of *S*s at different stages of their progress through a deck revealed that a strictly steady state was not completely achieved. In particular an analysis of the data from the fixed delay condition, presented in Fig. 2, clearly showed that the probability of classification of a new number as "old" (the probability of "false alarm"), rather than remaining constant throughout a deck, was still increasing after as many as 200 presentations. However, since the rate of increase of this probability after the first 40 cards was less than .0007 per card, the conditions of the experiment are regarded as "approaching a steady state."

The fact that the probability of classifying a new number as "old" increased with the number of other three- digit numbers already examined raises, again, the question of the effect of experience with previous

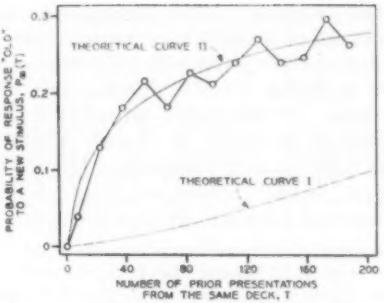

FIG. 2. Probability of classification of a new three-digit number as "old" as a function of the total number of three-digit numbers already seen in the same deck. (The points are averaged from all eight decks of the fixed delay condition.)

decks. If the trend illustrated in Fig. 2 had continued from one deck to the next (as well as within the same deck), the probability of classifying a new number as "old" should have increased as a function of the number of decks to which *S* had already been exposed. However the results for the fixed delay condition, which are summarized in Table 1, suggest that if the probability of the response "old" changed at all over the four successive decks it probably decreased rather than increased. (This trend, however, is not statistically significant.)

TABLE 1

EFFECT OF MEAN DELAY AND ORDER OF PRESENTATION OF A DECK UPON THE TWO TYPES OF ERROR PROBABILITIES FOR CARDS OF THAT DECK

Order of Presentation of Deck	Probability that New Number is Called "Old"					Probability that Old Number is Called "New"					Mean Probability for Errors of Both Types
	Mean Delay for Deck				Mean Probability	Mean Delay for Deck				Mean Probability	
	A 2.5	B 7.5	C 12.5	D 17.5		A 2.5	B 7.5	C 12.5	D 17.5		
1	.18	.21	.27	.29	.25	.15	.22	.24	.30	.23	.24
2	.17	.26	24	.30	.24	.17	.26	.25	.28	.24	.24
3	.14	.22	.26	.25	.22	.14	.27	.27	.35	.26	.24
4	.17	.18	.22	.22	.20	.17	.25	.30	.33	.27	.23
Mean	.17	.22	.25	.27	.23	.16	.25	.27	.32	.25	.24

The median 2-day separation of successive sessions may have sufficed, then, to reduce interdeck interference to negligible levels. Furthermore, of course, *Ss* may have adjusted their over-all tendency to classify stimuli as "old" so as to compensate for shifts that would otherwise have been observed from one deck to the next.

Discussion

The objectives of this section are to explore some implications of the data presented in Fig. 1 and 2 both for theories of forgetting and for estimates of the information capacity of human memory.

Implications for trace-decay models.— A stochastic model originally developed to account for the shape of the gradient of generalization (Shepard, 1958) has recently been extended to account for phenomena of recognition memory as well (Shepard, 1961). This extended model (referred to here as "the trace model") differs in two respects from previous models of the memory process. First, most previous models predict that the curve of forgetting is a simple exponential decay function with, perhaps, a positive asymptote (Bourne & Restle, 1959; Bower, 1959; Estes, 1955; Hull, Hovland, Ross, Hall, Perkins, & Fitch, 1940, pp. 43, 65; London, 1950; von Foerster, 1950; Witte, 1959; and, in a somewhat different form, Björkman, 1958, pp. 53–55). In contrast to these, the present trace model predicts a simple exponential only under the very special condition that all the presented stimuli are equally similar to each other. If the stimuli vary substantially in similarity, the predicted curve of forgetting departs from the best fitting simple exponential function in that it drops more rapidly at first and then more slowly than that function. Interestingly, then, many previously obtained empirical curves do appear to deviate from a simple exponential in just this way (see, for example, Hanawalt, 1937, p. 25; Jenkins & Dallenbach, 1924, p. 609; Luh, 1922, p. 22; Strong, 1913, p. 350; Williams, 1926, p. 373). That this is also the case for the present data is shown in Fig. 1. There, the trace model was fitted on the assumption that all three-digit numbers are equally similar. The resulting exponential decay function (Theoretical Curve I) clearly departs from the empirical points.

The second distinguishing feature of the present trace model is that it provides an account for the way in which the probability of "false alarm" increases as more and more stimuli are presented (as shown in Fig. 2). This follows from the assumption of the model that forgetting results not only from the fading or simple attrition of traces but also from the spreading out or "diffusion" of traces to other similar stimuli. The specific results of the attempt to fit this model to the present data will now be described.

The probability that a stimulus will be classified as "old" is denoted by $P_d(T)$. Figure 1, then, plots the mean values of $P_d(T)$ (i.e., $P_d(T)$ averaged over all positions in the deck from $T = 0$ through, say, $T = 200$) against delay, d. Figure 2 plots the probability of false alarm, which (according to the model) is given by $P_\infty(T)$, against the number of preceding presentations of stimuli, T. In order to fit the model, a value had to be determined for the response bias, B. Since *Ss* were not explicitly instructed to adopt either a stringent or lenient criterion in deciding whether each three-digit number was old, the initial expectation was that there would be no over-all bias. However, when B was set to unity (the null bias), even the most favorable setting of the remaining parameters failed to eliminate a clear departure from the data of Fig. 2 (see Theoretical Curve I). The discrepancy suggested that there was in fact an appreciable bias towards the response "old."

This was particularly evident when the over-all probability of responding "old" (regardless of whether the stimulus was in fact new or old) was plotted as a function of the mean delay for a deck. Theoretical Curve I in Fig. 3 represents the prediction from the model on the

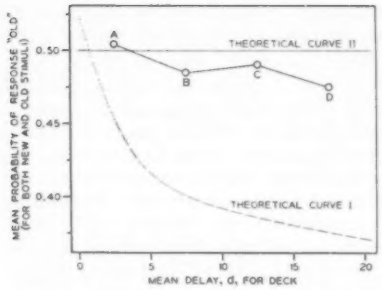

FIG. 3. Average probability of classification of both old and new three-digit numbers in a deck as "old" as a function of the mean delay between first and second presentations of numbers in that deck. (The point labeled "A" represents the mean probability for Decks A_1 and A_2 of the fixed delay condition, etc.)

assumption of no bias. But the empirical points, far from following that curve, remained more nearly in the proximity of the .5-probability line (Theoretical Curve II). There is a suggestion, therefore, that *S*s tended to readjust their response bias for each deck so as to match the objective probability of an old stimulus (which, as they could infer from the instructions, was always .5).

In accordance with this argument, when B was allowed to assume a value smaller than unity (representing a bias towards the response "old"), the theoretical curves were brought into a satisfactory fit with the data. The resulting estimates for the five parameters of the model were then as follows: $B = .42$, $w = .530$, $V = .996$, $V_e = .61$, and $u = .990$. If, as an approximation, the stimuli are regarded as falling in two classes; namely, one containing relatively distinctive three-digit numbers (such as the number 444 or *S*'s street address) and one containing relatively similar numbers (such as those in the cluster 355, 553, 335, and 533), then the parameters can be interpreted as follows (see Shepard, 1961): V represents the rate of diffusion of traces between the more distinctive numbers; V_e represents the rate of diffusion within clusters of similar numbers; w reflects the fraction of the stimuli that are more distinctive; and u specifies the rate at which traces dissipate through simple attrition.[3] When these values were substituted into the equations of the model, the curves labeled "Theoretical Curve II" in Fig. 1 and 2 were obtained. By adjusting $2\frac{1}{2}$ parameters per plot, then, the achieved fit was apparently about as good as the variability of the data would allow.

Implications for informational estimates.—An expression has already been derived for a lower bound on the amount of information retained in recognition experiments in which each stimulus is presented only once and in which steady state is reached so that the probabilities $P_d(T)$ no longer depend upon T (Shepard, 1961). In the present experiment each stimulus was presented twice and the probabilities $P_d(T)$ were still increasing at the end of 200 presentations. However the trace model provides for an extrapolation to the conditions required for the informational calculations. According to the empirically determined values of the parameters of the model, the steady-state curve of forgetting is described by the equation:

$$P_d(\infty) = \{.500(.986)^d + .443(.60)^d + .057\}^B$$

And, since the traces are assumed to dissipate and diffuse independently of each other, this same equation should also apply when each stimulus is presented twice (as in the present experiment). If the informational calculations are carried out on the basis of this equation with $B = .42$, the lower bound on the amount of information carried along by S during steady state is estimated as 32 bits. However, according to the model, *S*s could as well have adopted any other bias without influencing the

[3] By necessity, u must be less than unity; for, if there were no attrition of traces, S would eventually be swamped by the generalization from old stimuli and, thereafter, would be totally incapable of discriminating new stimuli from old. A parallel result emerged in the earlier application of the trace model to generalization in paired-associate learning (Shepard, 1958, p. 251).

underlying trace process wherein the information is actually retained. Therefore, since the informational estimate is affected by B, the possibility remains that *Ss* in fact carried along still more information that, owing to their response bias, was not completely reflected in their recorded responses. When B was varied in the steady-state equation, the informational estimate was found to climb rapidly from 0 (for $B = 0$), to level off at a maximum of about 50 bits near the null bias point ($B = 1$), and then to decline slowly with increasing B. This suggests that *Ss* are really capable of carrying along at least 50 bits in steady state. However, in order to secure this conclusion, it would be necessary either (*a*) to drive the parameter B away from the .42-region through the experimental introduction of various payoff contingencies or else (*b*) to replace the present "old"-"new" dichotomy with a scale accommodating several degrees of confidence that the stimulus is "old" (e.g., as was done by Egan, 1958).

In the familiar digit-span test *Ss* have generally been found capable of retaining about 23 bits from a single exposure to a sequence of decimal digits (Miller, 1956; Pollack, 1953). The present informational estimates, though based on the retention of decimal digits also, are substantially higher. Possibly the use of the more sensitive recognition method (Luh, 1922) and/or the segmentation of the sequence of digits into groups of three (Martin & Fernberger, 1929; Pollack et al., 1959) is responsible for this difference. Then too, the present procedure and the digit-span procedure may tap essentially different modes of operation of the memory system. Certainly the information is retained under the present procedure in a much more diffuse form. Whereas in the digit-span test each of a small number of items must be retained with high probability, here each of a very large number of items can be retained with relatively low probability.

The memory process has previously been described in terms of fading traces that may or may not be reinforced by covert rehearsal (e.g., by Brown, 1958; Mackworth, 1959; Sperling, 1960). However the experimental results, particularly those of Sperling and of Peterson and Peterson (1959), have indicated that the passive trace, if not maintained by rehearsal, decays essentially to zero within a matter of seconds. This seems to conflict with the present evidence for a substantial trace after several minutes (corresponding to 60 intervening presentations). However, although the rehearsal of one item following the presentation of a subsequent item probably seldom occurred, each three-digit number may well have been rehearsed once or twice before passing on to the next number. Such an initial boosting of the trace might account for its subsequent persistance over greater delays.

Summary

The retention of information was investigated under conditions approaching a steady state. Specifically, for each three-digit number in a protracted sequence of such numbers, *Ss* responded "old" or "new" depending upon whether they did or did not remember having seen that number at some earlier point in the sequence. Each number was presented twice; but the number of other three-digit numbers intervening between the first and second presentation of the same number was manipulated as the independent variable. The results were as follows:

1. The probability of correct classification of an old three-digit number as "old" was found to be a decreasing function of the number of intervening presentations. This "curve of forgetting" dropped (from near 1.00) to about .56 for delays of 60 intervening presentations. Even after such long delays, however, the probability of correct recognition was still well above the "false alarm" probability of classifying a *new* stimulus as "old."

2. This probability of false alarm was not constant over the series of numbers but, rather, climbed rapidly (from 0.0) at first and then seemed to be slowly approaching an asymptote near .30 after about 200 presentations.

3. A trace model was shown to yield theoretical curves that could be brought into a close fit with the above two curves. The fact that the curve of forgetting departed systematically from a simple exponential

decay function was interpretable within the model as indicating that the stimuli were not all equally similar to each other. The shape of the false-alarm curve could be interpreted as indicating that *S*s tended to bias their responses towards the response "old" in such a way as to match the objective probability of an old stimulus. Both curves support the notion that traces not only fade away but, at the same time, gradually spread out so as to become decreasingly stimulus specific.

4. The fitted curves also made possible the estimation of a lower bound on the amount of information that *S*s are able to carry along at any one time during an arbitrarily protracted sequence of three-digit numbers.

References

ANDERSON, N. S. Poststimulus cuing in immediate memory. *J. exp. Psychol.*, 1960, **60,** 216–221.

BJÖRKMAN, M. *Measurement of learning.* Stockholm: Almqvist & Wiksell, 1958.

BOURNE, L. E., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66,** 278–296.

BOWER, G. H. A theory of serial discrimination learning. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959. Pp. 76–93.

BROWN, J. Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, 1958, **10,** 12–21.

CAHILL, H., & HOVLAND, C. I. The role of memory in concept formation. *J. exp. Psychol.*, 1960, **59,** 137–144.

EGAN, J. P. Recognition memory and the operating characteristic. *USAF Cambridge Res. Cent. tech. Note,* 1958, No. AFCRC-TN-58-51.(AD-152650)

ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62,** 145–154.

HANAWALT, N. G. Memory trace for figures in recall and recognition. *Arch. Psychol., NY,* 1937, No. 216.

HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.

JENKINS, J. G., & DALLENBACH, K. M. Oblivescence during sleep and waking. *Amer. J. Psychol.*, 1924, **35,** 605–612.

LLOYD, K. E., REID, L. S., & FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60,** 201–207.

LONDON, I. D. An ideal equation derived for a class of forgetting curves. *Psychol. Rev.*, 1950, **57,** 295–302.

LUH, C. W. The conditions of retention. *Psychol. Mongr.*, 1922, **31**(3, Whole No. 142).

MACKWORTH, J. F. Paced memorizing in a continuous task. *J. exp. Psychol.*, 1959, **58,** 206–211.

MARTIN, P. R., & FERNBERGER, S. W Improvement in memory span. *Amer. J Psychol.*, 1929, **41,** 91–94.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63,** 81–97.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58,** 193–198.

POLLACK, I. The assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66,** 421–435.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57,** 137–146.

SHEPARD, R. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychol. Rev.*, 1958, **65,** 242–256.

SHEPARD, R. N. Application of a trace model to the retention of information in a recognition task. *Psychometrika,* 1961, **26,** 185–203.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

STRONG, E. K. The effect of time-interval upon recognition memory. *Psychol. Rev.*, 1913, **20,** 339–372.

VON FOERSTER, H. Quantum mechanical theory of memory. In H. von Foerster (Ed.), *Cybernetics: Transactions of the sixth conference.* New York: Josiah Macy, Jr. Foundation, 1950.

WILLIAMS, O. A study of the phenomenon of reminiscence. *J. exp. Psychol.*, 1926, **9,** 368–387.

WITTE, R. S. A stimulus-trace hypothesis for statistical learning theory. *J. exp. Psychol.*, 1959, **57,** 273–283.

YNTEMA, D. B., & MUESER, G. E. Remembering the present state of a number of variables. *J. exp. Psychol.*, 1960, **60,** 18–22.

(Received July 20, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 310–311

SUPPLEMENTARY REPORT: INFORMATION DEPRIVATION AND IRRELEVANT DRIVE AS DETERMINERS OF AN INSTRUMENTAL RESPONSE[1]

AUSTIN JONES

University of Pittsburgh

Some behavior, particularly of humans, appears interpretable as instrumental acts serving to reduce environmental uncertainty. In two experiments by Jones, Wilkinson, and Braden (1961), college *Ss* were deprived of visual information by requiring them to lie quietly in a darkened room. In Exp. I, *Ss* were allowed continuous use of a button which caused a series of lights to appear. Response frequency increased sharply over the first 9 hr., falling off slightly in the last 3 hr. In Exp. II, response frequency was an increasing function of the amount of relative information contained in the patterning of the light series. The experiments were interpreted as supporting a drive formulation of information deprivation.

The present experiment continues the assessment of an information drive concept by studying the effects of aversive and appetitive irrelevant drives upon the instrumental response of information-deprived *Ss*. Hunger, and the pain associated with electric shock were selected as the irrelevant drives. The hypotheses were derived from a view of drive summation consistent with that advanced by Hull (1943). It was hypothesized that increasing proportions of control *Ss*' responses would be made over time, with a slight decrement of response in the last quarter, as was found in the previous study, but that proportions of responses in the shock group would be facilitated in the initial hours of the session, then followed by a decrement over the remaining hours as a result of accommodation to the repeated presentations of the shock. (Such accommodation has been clearly shown in a conditioned avoidance procedure reported by MacDonald [1946].) It was hypothesized that *Ss* entering the session under 14 hr. food deprivation would have higher proportions of responses associated with the initial hours than would control *Ss*, and that proportions of responses would be an increasing monotonic function of hours, rather than decreasing in the last quarter, as predicted for control *Ss*. The hypothesis of an increasing monotonic function is necessitated by the assumption that the drive stimuli associated with food deprivation increase with hours of deprivation, at least over the moderate deprivation range employed here.

[1] This study was supported in part by Grant M-2479 from the National Institute of Mental Health, United States Public Health Service. The author is indebted to Phillip Davis for assistance in the execution of the experiments and to Harry Fowler for comments on the manuscript.

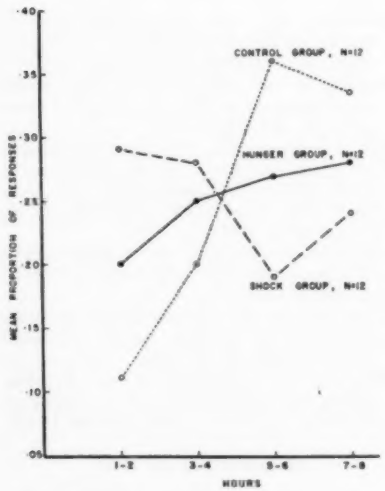

FIG. 1. Mean proportions of total responses as a function of hours in the deprivation situation.

Method.—The procedure was similar to that of the previous study. Thirty-six college *Ss* served in the information-deprivation condition for 8 hr. They were divided into three groups, one receiving occasional electric shocks throughout the session, another being deprived of food beginning 14 hr. prior to the session and continuing to its end, and a control group. The *Ss* were allowed to press a button causing patterned series of lights to appear, the relative information value of the series being determined by *S* through the setting of a dial.

Results.—The trends of response over time were compared by analysis of variance. The shock group and hunger group trends were each found to differ significantly from the control group trend ($P < .01$ and $P < .05$, respectively, two-tailed tests). As may be seen in Fig. 1, the differences are of the kind

hypothesized from the drive-summation formulation. The results are interpreted as further indication of the usefulness of an information drive concept.

A separate analysis of variance analyzed the trends of response over information values of the 36 *Ss* pooled. Response was shown to be an increasing linear function of the amount of information associated with the light series ($P < .025$), thus corroborating a principal finding of the previous study.

References

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

JONES, A., WILKINSON, H. J., & BRADEN, I. Information deprivation as a motivational variable. *J. exp. Psychol.*, 1961, **62**, 126–137.

MACDONALD, A. The effect of adaptation to the unconditioned stimulus upon the formation of conditioned avoidance responses. *J. exp. Psychol.*, 1946, **36**, 1–12.

(Received July 26, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 3, 311–312

SUPPLEMENTARY REPORT: THE INFLUENCE OF ONE STIMULUS ON THE PREDICTION OF THE ALTERNATIVE STIMULUS IN TWO-CHOICE PROBLEMS

IRVIN RUBINSTEIN

Walter Reed Army Institute of Research

Goodnow, Rubinstein, and Lubin (1960) have shown that *S* tends to follow the conditional probabilities of a stimulus in making his predictions. The concern of this report is the effect on *S* of the conditional probabilities of one stimulus when he is predicting after occurrences of the alternative stimulus.

Method.—Two groups of *Ss*, X and Y, were asked to predict a series of blue and orange cards. The length of the series for Group X was 124 trials, for Group Y it was 126. Group X was given a conventional 50:50 series randomized on both blue and orange. It was arranged so that all conditional probabilities were equal to .50 and there were no runs of either color greater than 5. For Group Y the series was 49% blue to 51% orange overall. The slight discrepancy from .50 was undesirable but unavoidable. The runs of blue for this group were exactly the same as those for Group X, not only in overall composition but also in the order in which the runs were presented. The orange cards were, however, always presented in run lengths of two for Group Y. There were 32 of these runs of orange. The conditional probabilities differed markedly from those of the orange runs given Group X.

Procedure was similar to that used by Goodnow et al. (1960). The *Ss* were United States Army enlisted men. Each group made up a special type unit in the United States Army. Consequently the two groups were well matched for job and rank. There were 38 *Ss* in Group X and 37 in Group Y.

Results.—The series has been split into thirds for the analysis. Since there were a total of 32 runs of each stimulus, the first third includes Runs 1–10, the second, Runs 11–21, and the third, Runs 22–32. Two methods of analysis appear feasible. First, we may compare the differences between the two groups on their prediction of blue after the occurrence of 1b (a single blue) and 2b (a double blue). This should be decisive since the only differences between their stimulus series were in the structure of the orange. Since $p(o/bo)$ for Group X is .50, while for Group Y it is 1, we expect that Group Y would be higher than Group X in its predic-

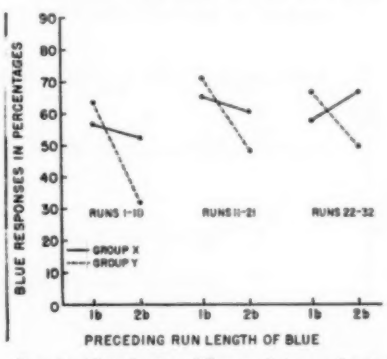

FIG. 1. Plot showing difference between the two groups in their choice of blue following one blue (1b) and two blues (2b).

tion of blue following 1b. On the other hand since for Group X p(o/boo) is .50, while for Group Y it is 0, Group Y should be lower than Group X in its prediction of blue after a run of 2b. The second approach is simply in taking the correlation in Group Y between *S*'s prediction of blue given 1b, p(b/ob), and his prediction of orange given 1o, p(O/bo), and also between p(B/obb) and p(O/boo)[1]. These conditional probabilities are unrelated insofar as the structure of the stimulus series is concerned. Any correlation would, therefore, have to be due to the tendency of *S* to react similarly to both stimuli irrespective of their differential structures.

Figure 1 shows the percentage of choices of blue for the two groups when the series is split into thirds. Since Group Y is expected to exceed Group X's prediction of blue after 1b, whereas the reverse should be true for the prediction of blue after 2b, the difference *D* between the two scores has been taken. This *D* score should maximize the effect of the experiment.

The t ratios between the *D* scores of the two groups for the three thirds are 3.86, 2.95, 3.67, with 73 *df*. The significance levels are .001, .01, and .001, respectively.

[1] Upper case letters refer to responses, lower case to stimuli.

In Group Y, $r_{(B/ob)(O/bo)}$ for the three thirds are .34 (P = .05), .35(.05), and .31(.10). The $r_{(B/obb)(O/boo)}$ are .37(.05), .10(NS), and .60(.001).

The hypothesis appears to be confirmed from both approaches. The t ratios given are all in the expected direction. The significance levels are given from the two-tailed test, although a one-tailed test would be justified.

Discussion.—The fact that *S*s in predicting two-choice series respond to the length of run of like symbols has been demonstrated by Goodnow et al. (1960). The results of the present experiment point out, however, that *S* is also influenced by the length of run of the alternative stimulus. The *S* appears to confuse and intermix his ideas about the two separate events. This appears to be so much the case, that aside from the statistical results given in this paper, only one of 37 *S*s was able to verbalize the situation given Group Y, where the run structures of the two colors were utterly different.

REFERENCE

GOODNOW, J. J., RUBINSTEIN, I., & LUBIN, A. Response to changing patterns of events. *Amer. J. Psychol.*, 1960, **73**, 56–67.

(Received January 17, 1961)

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland
PROF. DR. PH. LERSCH, München; Prof. DR. P. SANDER, Bonn

Herausgeber für Österreich
PROF. DR. H. ROHRACHER, Wien

Herausgeber für die Schweiz
PROF. DR. H. MEILI, Bern

Redaktion
DR. C. J. HOGREFE, Göttingen

Band VIII/1

HANS BÖTTCHER, Die Beurteilung der Prämorbidität bei Neurosen

HARDI FISCHER UND ALFRED LEDER, Untersuchungen über die Begriffsauffassung bei Jugendlichen

JOHANNES GRUNERT, Strukturspezifische Unfallkeime des Kraftfahrers

KURT HAUSS, Aufbau und Verlauf spezieller psychischer Funktionen im Bereich der menschlichen Gegenstandswahrnehmung

BERNHARD KRAAK, Räumliches Vorstellen als Voraussetzung technischen Denkens

GUSTAV A. LIENERT, Die Farbwahl im Farbpyramidentest unter Lysergsäure-diäthylamid (LSD)

WOLFGANG MEYER: Deutsches Autostereotyp und Heterostereotype von Amerikanern und Persern

ARTHUR B. SWENEY, Faktorenanalytische Methoden in der Motivationsforschung Cattells

WERNER TRAXEL, Über die Verwendung von Tonbandgerät und Schallschlüssel zur Zeitmessung sprachlicher Reaktionen

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.
Subscription is DM 64.—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
(Brentanoweg 10, Germany)

We take pleasure in announcing . . .

New Executive Editor

Our new executive editor is John J. Boren, The Institute for Behavioral Research, University of Maryland, College Park, Maryland. Please direct future manuscripts to this address.

New Apparatus Editor

Our new apparatus editor is N. H. Azrin, Behavior Research Laboratory, Anna State Hospital, Anna, Illinois. Please submit apparatus notes directly to this address.

New Consulting Editors

To process an increasing number of manuscripts, we have added six new consulting editors to our board.

New Format

To increase the number of articles we can carry in each issue, we have adopted a new space-saving double-column format.

Journal of the Experimental Analysis of Behavior

Back volumes and regular subscriptions: $12.00 per year. Special rates for current subscriptions: members of recognized professional societies, $9.50; students, $6.50. All subscriptions for a given year begin with the January issue. Please continue to address all subscription and business correspondence to:

Kay Dinsmoor, JEAB
Psychology Department
Lindley Hall
Indiana University
Bloomington, Indiana